Reinforcers and Control

Towards a computational ætiology of depression

Quentin JM Huys

Dissertation submitted for the degree of
Doctor of Philosophy
of the
University of London

Gatsby Computational Neuroscience Unit
University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

I, Quentin Jan Marie Huys, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

June 29, 2007

ABSTRACT

Depression, like many psychiatric disorders, is a disorder of affect. Over the past decades, a large number of affective issues in depression have been characterised, both in human experiments and animal models of the disorder. Over the same period, experimental neuroscience, helped by computational theories such as reinforcement learning, has provided detailed descriptions of the psychology and neurobiology of affective decision making. Here, we attempt to harvest the advances in the understanding of the brain's normal dealings with rewards and punishments to dissect out and define more clearly the components that make up depression. We start by exploring changes to primary reinforcer sensitivity in the learned helplessness animal models of depression. Then, a detailed formalisation of control in a goal-directed decision making framework is presented and related to animal and human data. Finally, we show how serotonin's joint involvement in reporting negative values and inhibiting actions may explain some aspects of its involvement in depression. Throughout, aspects of depression are seen as emerging from normal affective function and reinforcement learning, and we thus conclude that computational descriptions of normal affective function provide one possible avenue by which to define an ætiology of depression.

CONTENTS

List of Figures

LIST OF TABLES

PREFACE

ACKNOWLEDGEMENTS

Throughout the four happy years I spent at the Gatsby unit, Peter Dayan, my supervisor, has given me continued, constant, patient and extraordinarily insightful guidance and advice that went far, far beyond the call of duty. To acknowledge the extent to which he was involved in every aspect of the work presented here, I will throughout the thesis use the first person plural rather than singular. However, my debt towards him extends further, as I owe him much of my view on science and the theoretical foundations to pursue the kind of research I have come to believe in. I would also like to thank Peter Latham for the patient support early on during my time at the unit.

My parents, my sister and my friends, particularly Sofia Russo, have been extremely tolerant of the demands that doing a PhD put upon me. Without them, none of this would have been possible. Throughout the time at the Gatsby, the people there have been an immense source of inspiration and strength, and I am very grateful to have come to know each one of them.

I am indebted to Jonathan Williams for getting me interested in psychiatry, for making me believe that computational neuroscience can indeed be usefully applied to psychiatric issues, and for the invaluable input throughout this project. Similarly, I am also very grateful to UCL's MB/PhD programme and its directors, Gordon Stewart and Anthony Segal, for allowing me to embark on a PhD that was so ostentaciously remote to medicine.

For reading parts or all of the thesis and providing so many important inputs I would like to express my sincere appreciation to Nathaniel Daw, Peter Dayan, Máté Lengyel, Michael Moutoussis, Hanneke Den Ouden, Emma Sweeney, Jonathan Williams and Kimberley Wilson. Great thanks are also due to the many people who provided feedback on this work in its various embryonic stages, particularly Ian Anderson, Ryan Bogdan, Y-Lan Boureau, John Christianson, Nathaniel Daw, Bill Deakin, Ray Dolan, Máté Lengyel, Michael Moutoussis, Stephen Maier, Yael Niv, Diego Pizzagalli, J Douglas Steele and last but most certainly not least Paul Willner, who has influenced much of this thesis both directly during my visit in Swansea and through his writing. I would also like to thank Ray Dolan and his emotion club for feedback when I gave a talk at the FIL, particularly Hanneke Den Ouden for many fun hours of discussion.

Finally, I would like to thank Barbara Sahakian and Karl Friston, who kindly agreed to serve as examiners for this thesis and gave very valuable feedback.

OTHER WORK DURING THE PHD

This thesis is the result of the last year of my time at the Gatsby Unit, and none of this work is published as yet. Before approaching depression, I worked on other issues, which have resulted in the following publications:

- **Population coding in time** (Zemel et al., 2005; Huys et al., 2007; Natarajan et al., 2007)
- **Efficient estimation of detailed biophysical models of single neurones** (Ahrens et al., 2006; Huys et al., 2006; Huys and Paninski, 2006)
- **Gaussian process applicatons to medical classification problems** (Nouraei et al., 2007) **and to the inference of smooth intensity functions for point process observations (Minor project in machine learning;** Huys 2006).

To my family and my London friends.

I

INTRODUCTION

Psychiatry has come far over the past century. Arguably the main — and most controversial — advance is the emergence of a tentative nosological agreement amongst practitioners. This consensus has carried in its wake ever more detailed and sophisticated explorations of the psychological, neurobiological and social nature of psychiatric diseases, which closely mirror the great progress made in our understanding of the healthy brain. We have now reached an exciting point: the various approaches have not only matured enough to be brought together in a thorough normative framework but the framework is already in existence. A central goal of this thesis is to argue that affective decision making, broadly defined as decision making or action choice in the context of reinforcers, provides a natural frame of reference for psychiatry. Our understanding of normal affective decision making has been formalised by computational neuroscience over the past two decades. It forges strong links between precisely those aspects of psychology and neurobiology that are classically disturbed in psychiatric illness. It allows us not only to gain a deeper understanding of why the neurobiological changes lead to the psychological and behavioural constellations found in clinical practice but it can also serve a guiding role for further, theoretically motivated investigations. Furthermore, the link between normal function and disorder promises to provide the so-far elusive definition of mental disease, as opposed to the descriptive syndromes to which we have become so used.

1.1 DEPRESSION

In this thesis, we will attempt to apply this approach to one particular mental illness: depression. The discrepancy between the intuitive clarity and familiarity[1] of the concept and the difficulty commonly encountered when looking for a definition of depression that is more than

[1] Melancholia ($\mu\epsilon\lambda\alpha\gamma\chi o\lambda\iota\alpha$) is the only condition from the Hippocratic classification to have survived to today (Wong and Licinio, 2001).

mere tautology is baffling. Just as baffling, we are told by those who have actually suffered from it is the abyss between the common conception of depression as a state of low mood and the blackness that is depression.

> [Depression is a word] that has slithered through the language like a slug, leaving little trace of its intrinsic malevolence and preventing by its very insipidity, a general awareness of the horrible intensity of the disease when out of control. (Styron, 1991)

> David had died. [...] But grief, fortunately, is very different from depression: it is sad, awful, but it is not without hope. (Jamison, 1997)

Unfortunately, our understanding of depression is still very limited.

> "If you compare our knowledge [of depression] with Columbus' discovery of America, America is yet unknown; we are still down on that little island in the Bahamas" (Styron, 1991)

Ideas about the nature of depression are abundant, as are ideas about what it might mean to understand depression. However, most people agree that depression is a disorder of affect, and the main classes of theories — biological, behavioural and cognitive — describe different, but interrelated, facets of the affective changes. It is our contention that affective decision making provides an integrative framework for these theories. Let us briefly introduce the reasoning and then formulate the questions that will guide our review of the literature and our investigations.

Biological theories postulate that depression is an organic disorder of the brain. They originate from the existence of effective pharmacological therapies and are more recently being merged with genomic approaches (Licinio and Wong, 2004). In their purest forms, they make no claims at all about the psychological processes involved. The clearest result from this field is that depression (as indeed most psychiatric disorders) is related to dysfunctions in neuromodulatory systems, mainly serotonin, but also dopamine (DA) and noradrenaline.

Unadultered behavioural theories make no claims at all about the biology involved, concentrating instead on providing consistent descriptions of observable behaviours. Behaviourists focus on "conceptual nervous systems" (Hebb, 1955), which describe how stimuli, be they internal or external, lead to particular actions. In terms of depression, the stimuli are affective ones: rewards and punishments. The main claims are that depression is related to 1) a decreased efficacy / frequency of reinforcers; 2) the perception that reinforcers are not "controllable" and 3) that depression may be induced by stressors (Blaney, 1977).

Cognitive researchers emphasise the primacy of subjective reality. The approach has taken much inspiration from the behavioural work and in essence addresses similar affective issues, but couched in quintessentially social, human terms (Beck, 1967; Seligman, 1975; Abramson et al., 1989; Lewinsohn et al., 1979). Their main "instruments" are interviews, in which subjects are expressly asked about their feelings and thoughts. Practitioners do not attempt to ascertain the true nature of events, but concentrate on the interpretation individuals give to these events. The main achievement of this approach is a translation of the behavioural concepts into non-pharmacological treatments for mild depression, showing that remission from depression can

be achieved by "removing" stressors, e.g. by teaching individuals how to cope with them or how to re-interpret events in less stressful manners.

Our aim is to show that affective decision making can integrate these three facets because

1. it gives neuromodulators and stressors precisely circumscribed roles in behaviour, thereby linking biological and behavioural theories
2. it mathematically formalises the concepts underlying the cognitive theories. These formalisations make clear behavioural predictions, and thus link cognitive theories to both behavioural and biological ones.

It is hoped that such an integration will allow a clearer understanding both of the manifold ways to depression (how alterations to different parts of the involved systems can lead to similar depressed states) and the relationship between different states all characteristic of depression (in the current predominant diagnostic framework (DSM-IV), it is possible for two depressed patients to share no symptoms).

1.2 NEUROBIOLOGY OF AFFECTIVE DECISION MAKING

We now give a very brief overview of affective decision making. Our knowledge of affective decisions comes mainly from behavioural tasks in which subjects learn about reinforcers in order to guide action choice such as to maximise the rewards earned and minimise the costs incurred. At least four computational aspects of such problems are known to be relevant in animals, and have at least a partially explored neurobiological basis: Pavlovian, habitual and goal-directed action choice and motivation. In this thesis we will explore the relationship of these to depression, and suggest that features of depression arises from the interplay between them. As we will see in chapter 2, there is evidence that depression affects the primary sensitivity to reinforcers, and also each of the decision making systems.

Figure 1.1 shows an archetypal affective decision-making paradigm. A sequence of actions, here navigational actions through rooms of a house, has to be chosen such as to maximise the achieved *total* reward and minimise the total punishment. The difficulty comes from the fact that it is the total, rather than the immediate reward, that should be maximised. The three decision making systems we will discuss here are: goal-directed, habitual and Pavlovian decision-making systems, which vary in terms of the computational load, and the efficiency with which they use data. We will also discuss the effects of motivation on these systems.

Conceptually most straightforward is the *goal-directed* system which relies on a tree search. It assumes the consequences of all actions are known, simulates the consequences of all actions and outcomes, and chooses the best action sequence (figure 1.1A). Going through all the options can be thought of as a reflective, sequential planning process. It generalises directly to the case where actions have probabilistic outcomes, in which case action sequences are chosen according to the probabilities they afford to large rewards. But it is only feasible for small trees, i.e. when there is a small number of actions and outcomes, as the number of sequences (hereafter "paths") grows as $D^{|A||O|}$ (where D is the depth of the tree (the number of actions

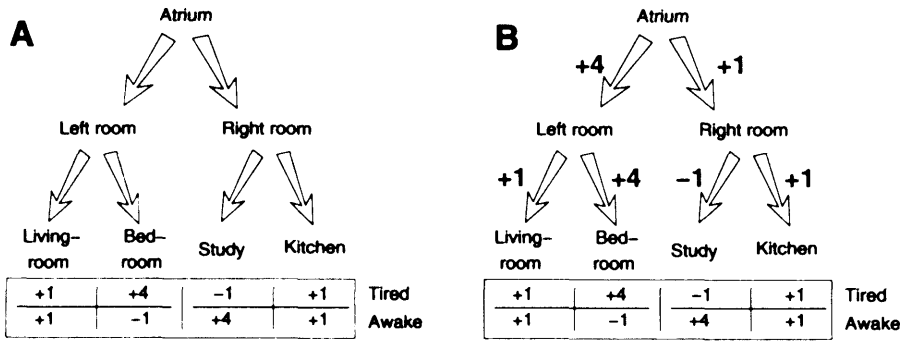


FIGURE 1.1: Basic decision making setup. **A**: Make the (admittedly unrealistic) assumption that a student in London has a labyrinthine house with seven rooms. Each room has an object in it which makes it more or less desirable when he is awake or tired. Before knowing the layout of the house well, he will have to try out the various options a least once. Having tried them all out, he knows the outcomes of each, can think through all options (plan) and choose the best sequence of actions given his state. If he is awake, right, then left, leads him to desirable books on depression in the study (+4 rewards). If he is tired, these books are aversive (-1 rewards), and he will choose an action sequence which leads him to the now desirable bedroom (+4 rewards). This is goal-directed action choice. It relies on a model of the world (what actions lead him where), and on a search through the whole tree of action-outcome sequences. **B**: One year year down the line the student has experienced each action in each room extensively when tired (he spent his awake time in the lab). He now knows what future rewards to expect from each action in each room, and habitually follows whichever on average yielded the best results. He knows the long-term outcomes of each action in each room and does not need to think through all options any more. Assume some day the computer server in the lab is down. He comes home, but is not tired. However, he still habitually chooses left, then right, only to find out that he actually does not want to sleep. So he back-traces and goes to the study with the books. Habitual action choice is not sensitive to the outcomes of actions. If a motivational shift occurs (here from tired to awake), the values of each action in each state (room) have to be learned anew.

to choose), $|A|$ the number of available actions at each choice, and $|\mathcal{O}|$ the number of outcomes after each action). It is more efficient, compared to the other methods below, if little is known about the rewards available (Sutton and Barto, 1998; Bertsekas and Tsitsiklis, 1996; Daw et al., 2005; Lengyel and Dayan, 2007). Cognitive theories of depression, and some behavioural theories, effectively argue that depression is characterised by alterations to this process (Maier and Seligman, 1976; Beck, 1987). Depressed people, the argument goes, wrongly estimate the likelihood that actions will evoke reinforcing outcomes (Lewinsohn et al., 1979; Layne, 1980; Blaney, 1977; Maier et al., 2006). Some additional support for such hypotheses comes from imaging studies, which rather consistently find abnormalities in depression in prefrontal brain areas associated with planning and executive function (Mayberg, 1997; Drevets et al., 1997). Appendix A.2 gives more details on tree search, section 2.4 will review the human data on its relationship to depression, and chapter 4 formalises the notion of control in goal-directed action choice.

Habitual action choice is computationally much less demanding than tree search, but it is only reliable after extended experience of an environment. Rather than recomputing the total expected future reward (the value $Q(s,a)$) for each action a in each state s, as above, this quantity is accumulated over repeated runs through the environment. Once known, choosing the optimal sequence of actions simply consists of choosing, in each state s independently, the action a that maximises the cached value $Q(s,a)$, i.e. $a_{opt} \equiv \arg\max_a Q(s,a)$. The final cached Q values are shown in red in figure 1.1B. Because they are constructed by averaging over many runs, Q values are insensitive to rapid changes, such as due to a sudden motivational shift which changes the reinforcing properties of outcomes (e.g. from tired to awake, rendering the bed aversive and the books attractive). The insensitivity to sudden shifts of outcome value is characteristic of habits. Temporal difference (TD) models, and their simpler counterparts such as the Δ rule (Rescorla and Wagner, 1972), are update equations that allow online estimation of Q values (Sutton and Barto, 1998; Watkins and Dayan, 1992; Bertsekas and Tsitsiklis, 1996). The similarity between the phasic activity of ventral tegmental area (VTA) dopaminergic cells during the acquisition of habits (or stimulus-response mappings) and the predictive error of the TD equations (Montague et al., 1996; Schultz et al., 1997; Schultz, 1998; Hollerman et al., 2000; Waelti et al., 2001; McClure et al., 2003a; Bayer and Glimcher, 2005) are strong indicators that phasic dopamine is involved in building predictions of rewards (incentive value; Bindra 1978; Berridge and Robinson 1998; McClure et al. 2003b) and the acquisition of habits (Nelson and Killcross, 2006)). Activity in a variety of limbic areas (e.g. striatum, orbital and medial prefrontal cortices, amygdala) correlates both with the predictive error signal and with the resulting value signal (McClure et al., 2003a; Gottfried et al., 2003; O'Doherty et al., 2001, 2003, 2004). Finally, we should point out that there is very good evidence that the brain uses both cached and tree-based approaches simultaneously (Dickinson and Balleine, 2002; Killcross and Coutureau, 2003; Daw et al., 2005; Dayan et al., 2006). Habitual action choice is implicated in depression both by decreases in DA metabolism (though there is no specific evidence for the involvement of phasic as opposed to tonic signals, see section 2.1.4) and by effects in behavioural paradigms (Richards and Ruff, 1989; Henriques et al., 1994; Pizzagalli et al., 2005). The claim is here that depressed subjects are less efficient at acquiring habits that have rewarding consequences, although we will see in the next chapter that this is more likely due to changes in the primary sensitivity to reinforcers.

The *Pavlovian* system is computationally the simplest and chooses an evolutionarily predefined set of actions based purely on valence information: simplistically speaking, rewards are approached, and punishments avoided. The strength of the reflexive appetitive approach actions, and also the fact that they may be non-adaptive in certain situations (Breland and Breland, 1961; Dayan et al., 2006), is exemplified by phenomena such as negative automaintenance, where pigeons are unable to suppress pecking a light predictive of reward, although this leads to reward omission (Holland, 1979), and also by experiments such as that by Hershberger (1986), in which a food tray receded at twice the animals speed when approached, but come towards the animal at twice its speed when the animal moved away from it. Animals were unable to learn to run away from the food and overcome the predominant appetitive approach behaviour. Aversive Pavlovian actions are richer than their appetitive counterparts and take more detailed facets of the aversive stimulus into account, such as its proximity (Blanchard and Blanchard, 1988), and the various actions are organised topologically in the peri-aquæductal grey (Bandler and Shipley, 1994). It is the control over the periaquæductal grey (PAG) by the serotonergic system and its simultaneous involvement in the prediction of punishments (see below) which is of interest to us and will be analysed in detail in chapter 5.

Motivation finally interacts with these systems to energise actions. For example, animals that are thirsty and sated will press both levers associated with food and those with water more vigorously. It energizes behaviour not only in this general manner, but also specifically with respect to appetitive outcomes, as it increases responding on a lever on which food is available more (Fletcher, 1996; Berridge and Robinson, 1998), and it affects the trade-off between rewards and efforts (Salamone and Correa, 2002; Phillips et al., 2003). Importantly, reports that it is controlled independently of the phasic signal (Floresco et al., 2003; Goto and Grace, 2005) suggest that it is indeed an independent valuation system. Recently, Niv et al. (2005, 2007) have united the various proposals in an average-reward (Mahadevan, 1996) extension of the TD learning framework, where animals choose both which action to emit, and when. Tonic levels of dopamine here are suggested to report the average level of reward expected for the emission of actions. The higher DA, the more it is worth emitting actions. However, it is at present unclear by what connections tonic DA comes to represent this information (Yael Niv, pers. comm.). We will see in chapter 4 that a similar notion of DA may have a role in the tree search, though once again the neurobiology underlying this is murky at best. Not only has DA itself been suggested to be abnormal in depression, but behavioural effects of tonic dopaminergic signalling have been suggested as core phenomena in depression (see chapter 2; Kapur and Mann 1992; Parker and Hadzi-Pavlovic 1996; Elliott et al. 1998; Austin et al. 1999, 2001; Willner 2002).

1.2.1 DOPAMINE AND SEROTONIN

Dopamine is closely involved with rewards, but there is also need for a system that reports punishments. DA neurones have low background firing rates, which limits their ability to signal negative prediction errors (due to either punishments or reward omissions). Patients with Parkinson's Disease have difficulties learning from positive reinforcements, compared to controls, but do not present such difficulties for negative reinforcements (Frank et al., 2004). This

mirrors the behaviour of DA neurones, which respond more to rewards, than they are suppressed by punishments or reward omissions (Schultz and Dickinson, 2000; Daw et al., 2002; Bayer and Glimcher, 2005). Indeed, there is direct evidence that the brain carves up the reinforcement signal into at least two opponent components. In trans-reinforcer blocking (Ganesan and Pearce, 1988), animals can be prevented from acquiring a response to a conditioned stimulus (CS) predictive of a) reward omission by simultaneous presentation of a CS associated with punishment; and b) one class of reward by presentation of a CS associated with another class of reward (e.g. food and water for animals that are both hungry and thirsty). Behaviourally, aversive learning provides a near perfect mirror of appetitive learning and on its own appears as well-described by reinforcement learning models as appetitive conditioning is (Bouton, 2006).

However, there is no dopamine of the aversive system. Serotonin (or 5-hydroxy-tryptamine, 5HT) has been suggested (Daw et al., 2002), but, unlike dopamine, it has not yet been convincingly shown to provide a "negative TD" signal: so far, physiological recordings of serotonin neurons have shown little more than an anticorrelation with arousal (Jacobs and Fornal, 1997), and some sensory events (Ranade and Mainen, 2006). This may of course be due to the much more complex and widespread pharmacology of the 5HT system. Nevertheless, many facets of these findings are important. In opposition to DA, it generally inhibits actions and has thus been associated with a behavioural inhibitory system (Wise et al., 1972; Carter and Pycock, 1978; Soubrié, 1986; Gray, 1991). Two distinct modes of action inhibition is seen in animals. Firstly, in promoting sensitivity to punishments, in that 5HT antagonists release punished responding and in that 5HT depletion impairs the withholding of actions and potentiates escape responses from aversive PAG stimulation (Deakin and Graeff, 1991; Fletcher, 1995; Graeff, 2002; Maier and Watkins, 2005). Secondly, in reducing sensitivity to rewards, in that 5HT infusions into the nucleus accumbens (NAcc) impair consummatory and preparatory responses for rewards, whereas pharmacological lesions potentiate it (Amit et al., 1991; Fletcher and Korth, 1999; Fletcher et al., 1999; Higgins and Fletcher, 2003). Thus, it appears that serotonin reports aversive events and prevents actions that lead to aversive outcomes. A priori it then seems odd to associate depression with decreased serotonin function, and we will concentrate on this conundrum in chapter 5. We will argue that this is an example of Pavlovian action choice, where actions are chosen in a reflexive manner based purely on reinforcer predictions.

If depression is, as biological theories have long suggested, related to malfunctioning or inefficient dopaminergic or serotonergic systems (Willner, 1985b), then it should be possible to use the normative data on these systems to devise specific behavioural tasks in humans to probe particular aspects of the systems in depression. Not only does this provide an aetiological framework, but it also promises to allow us a much more detailed dissection of depressive phenomena, potentially uncovering subtypes of the disorder that are coherent at a behavioural, neurobiological and cognitive level.

1.3 MODELLING

A few comments about modelling in general and in psychiatry are appropriate. Modelling efforts classically seek explanations at one of at least three levels (Marr, 1982; Dayan and Ab-

bott, 2001): what, how and why. What questions are addressed by mechanistic models and aim to summarise, concisely yet in detail, large amounts of data. How questions are addressed by mechanistic models, and aim to answer for example how features of neural circuits are explained by features of subcomponents, or how particular computations are implemented by neural hardware. Questions about why finally are addressed by interpretative models. Here, the behaviour of neural systems and organisms is explained with reference to certain principles, often principles of optimality (such Bayesian inference and reinforcement maximisation).

The main body of the thesis presents three modelling studies which primarily provide a description of behaviour in a normative setting. Normativity expresses the fundamental tenet that behaviour reflects optimal action choice, optimal in a Bayesian and reinforcement-learning sense: organisms are viewed as selecting actions such as to maximise their expected, long-term, reward given a) their knowledge about environmental action-reward structures, and b) constraints on the kinds of computations they are able to perform. Chapters 3 and 4 show how behaviour that has been related to depression can be viewed as optimal in the circumstances provided by the tasks in which it is observed. Where it is feasible, attempts are made to relate the computational building blocks of the models to particular functional aspects of neurobiology.

Therefore, these models provide one aetiological link between what is termed healthy and disordered *function*. However, we emphasize that they are abstract (non-physiological) models that aim to elucidate the function of the involved neurobiological components, not the particular way in which these functions are implemented by nervous tissue (although the latter is an important objective for future work). Thus, they are models situated at the "why" level, and the aim is to reproduce qualitative aspects of the data. The qualitative behaviour of the models reported throughout is only mildly sensitive to parameter choice.

Each of the three main chapters concentrates on one system, or at most on the interaction between two systems, although of course the observed behaviour is a product of the interaction of all the systems described in the previous section (and probably even more). It is hoped that exploration of the effects of few systems will pave the way for a better understanding of the full decision making process in depression (as outlined in chapter 6). It should be emphasised that a multitude of options exist for the interaction of these systems to yield non-optimal behaviour (Dayan et al., 2006).

Finally, while this work is in its early stages and not reported here, the models form the basis of and rationale for behavioural tasks in humans, the aim of which it is to ascertain specifically and directly whether depressed subjects differ on tasks that rely on any one of the decision making aspects outlined in the previous section. In particular, chapter 4 forms the basis of tasks to probe and infer the prior beliefs subjects hold about control; chapter 5 motivates investigations into tasks that probe Pavlovian effects in depressed subjects; and the generalisation issues discussed in chapters 3 and 4 lead directly to animal experiments, particularly in the light of recent data on the involvement of the hippocampus (Santarelli et al., 2003).

1.4 ORGANISATION OF THE THESIS

Chapter 2 first reviews the literature on the state of depression from a computational perspective. It is organised in terms of the affective decision making systems just outlined and addresses questions along these lines:

1. Primary
   Is there a change in primary reinforcer sensitivity? Is it confined to either valence, or symmetrical?
2. Pavlovian
   Is there evidence for altered approach or avoidance behaviour?
3. Goal-directed
   Do depressed people make different predictions about the likelihood of action outcomes and does this affect their goal-directed action choice?
4. Habitual
   Is there a change in the acquisition / expression of appetitive or aversive habits?
5. Motivation
   What motivational changes are there? How do these correlate with neuromodulator changes?

The chapter then proceeds to review the literature on the ætiology of depression and on the induction by stress of a state reminiscent of depression in animals.

The main body of the thesis presents the computational consequences of data on these aspects of decision making for the induction (hence ætiology) and maintenance of the disease. To isolate the effects as much as possible, we will focus on data that is not confounded by issues of consciousness. While we will mention some of it, we will not rely on data involving verbal or otherwise conscious reports. Thus, we will mainly base our arguments on human behavioural data to characterise the *state* of depression, and on animal data to gain insights into the *induction* of depression.

In chapter 3 we build a habitual model of learned helplessness. Our goal is to ascertain whether a change in primary reinforcer strength (analgesia), induced by stressors, can account for the data in animal experiments on learned helplessness. We find that it accounts for most of the data, but that several doubts persist, particularly in terms of how the effects generalise across reinforcer valence.

Chapter 4 also works on the induction of depression. We there give a theoretical formulation of control, explore its characteristics and relate it to the literature on depression.

Having discussed the induction of depression by stress, chapter 5 looks at aspects of the maintenance that may be due to serotonin's effect on both Pavlovian and instrumental systems. This sheds light on the confusing fact that serotonin is associated with punishment, yet drugs that increase serotonin provide relief in depression.

II

LITERATURE REVIEW: AFFECTIVE DECISIONS IN DEPRESSION

ABSTRACT

This section will first give a concise and focussed overview of the general features of depression as a disorder, its epidemiology, diagnosis and treatment. Thereafter, we review in some detail the evidence that implicates each of the facets of affective decision making outlined in chapter 1. There is evidence that the state of depression is associated with decreased sensitivity to reinforcers of positive and negative valence which have consequences for habit acquisition. The goal-directed system is mainly implicated in conscious reports in humans, but has not been sufficiently assessed behaviourally. Behavioural data and data on tonic levels of dopamine indicate that a subgroup of patients may suffer from decreased motivation. Serotonergic data finally argues for an involvement of Pavlovian mechanisms. Finally, data on the ætiology in humans and on animal models is reviewed. There is clear evidence that stress may have a causal role in the onset of depression and this is paralleled by the induction of a state reminiscent of depression in animals by stress.

The term depression applies to transient states of negative affect experienced by most people at some point in their lives. It also applies to a debilitating condition with severe health implications termed major depressive disorder (MDD, Diagnostic and Statistical Manual IV; American Psychiatric Association 1994), which can additionally include impairments of cognition, neurovegetative and motor function (Fava and Kendler, 2000). Henceforth, we will use depression to refer to MDD. After a short epidemiological overview over the scale of the prob-

lem depression poses to humanity, we proceed to clarify a few fundamental notions. The main aim of the literature review is to ascertain to what degree there is evidence for the involvement in depression of each of the facets of affective decision making described in section 1.2. We wish to emphasize three caveats at the onset: firstly, this knowledge is relatively circumstantial as it emerges from research of which it was not the primary concern. Secondly, we have no knowledge at all about the interrelationship between these changes. We do not know whether depression tends to affect many of these systems in any one individual or whether there is a temporal pattern to it. Thirdly, there are major confounds due to variability in the assignment of diagnoses and their inherent lack of specificity. Due to the sparsity of the behavioural data, we can at present only acknowledge these issues.

2.1 BACKGROUND

2.1.1 EPIDEMIOLOGY

> Il n'y a qu'un problème philosophique vraiment sérieux: c'est le suicide.
> [There is only one problem of real philosophical significance: suicide]
> (Camus, 1942)

By 2020, depression will be the major inducer of disability worldwide (World Health Organization, 1996). It correlates negatively with nearly all measures of quality of life (Trompenaars et al., 2006), and is associated with significant mortality, being the leading cause of suicide. Suicide kills 1,000,000 people every year across the world and over the past 20 years has killed 200,000 more US Americans than AIDS. Its deaths outnumber by a factor of four those who fell in the Vietnam war and by a third those who die in homicidal acts (Goldsmith et al., 2003). In comparison with the huge judicial apparatus designed to deal with homicide, the effort to deal with suicide is minor. A psychiatric cause, including substance abuse, is present in 90% of suicides in the western world. Depression is present in at least 50% of suicidal adults and nearly 80% of suicidal children (Goldsmith et al., 2003). Approximately 60-70% of acutely depressed patients experience suicidal ideation and some 10% actually commit suicide (Wong and Licinio, 2001), and measures of aspects of depression can be used to predict suicide (hopelessness; 91% sensitivity, 46% specificity over 10 years; Beck et al. 1989, 1990).

Depression is extensively comorbid with other psychiatric, neurological and general medical disorders (Wong and Licinio, 2001; Cryan and Holmes, 2005; Ninan and Cummins, 2003; O'Brien, 2006), particularly in at-risk populations (Power, 2005). Of most interest to the present work will be the strong link with anxiety disorders (Clark and Watson, 1991; Kessler, 1997; Mineka et al., 1998). 50-60% of patients with a lifetime history of major depressive disorder (see below) report a lifetime history of an anxiety disorder (particularly panic disorder and generalised anxiety (GAD)), usually with the anxiety disorder predating depression (Kaufman and Charney, 2000). Depression is more severe when comorbid with anxiety and has more severe consequences, e.g. in terms of suicide (Kaufman and Charney, 2000). GAD and MDD appear to fully share genetic contributions, with the differential development of the disorders dependent on the precise pattern of life events experienced (Kendler et al., 1992; Manicavasagar

et al., 1998; Mineka et al., 1998; Kaufman and Charney, 2000; Hettema et al., 2006a).

Despite important cultural variations (Bentall, 2003; Kleinman, 2004), depression extends across the world — the common cold of psychiatry. One-year prevalence rates in Europe vary from 1.4% in rural Bavaria to a staggering 53% in Japanese first-year university students (though some of this may be due to variation in diagnostic criteria). Lifetime rates range from 1.5% in Taiwan to 19% in Beirut, and throughout women are more at risk than men (average ratio of 2.5:1) (Weissman et al., 1996; Wong and Licinio, 2001; Bandelow, 2003).

Nearly one in five US Americans will experience depression at some point in their life (National Comorbidity Survey, ;Blazer et al. 1994), and most will have prolonged residual syndromes and relapses (Judd et al., 1998). Depression often comes early: a substantial proportion of people experience their first episode of major depression in childhood or early teens (Fava and Kendler, 2000), the average onset age is in the late 20s (Weissman et al., 1996), and the peak 12-month prevalence is in the late teens for women and the early 20s for men (Costello et al., 2002). Early onset tends to be predictive of more severe and recurrent depression (Weissman et al., 1999b,a; Lewinsohn et al., 1999).

Two main conclusions can be drawn from these facts: Firstly, it is so common that it is most likely informative about normal brain function. The flip side of this statement is that insights into depression should be available from an understanding of normal brain function. Secondly, depression is a major issue which deserves the massive scientific attention it gets.

2.1.2 Diagnosis

Advances in treatment and understanding of putative diseases are only possible when there is agreement over the entity under scrutiny. Arguably the main advance in psychiatric theory over the past two decades has been the emergence of a partial nosological consensus with the formulation of DSM-III. However, the classification of psychiatric diseases according to this consensus is syndromal and descriptive. In the main it is *not* driven by any understanding of the underlying causes of the disease (Hasler et al., 2004) or its treatment (Spitzer, 1998). Given the absence of an understanding of psychiatric diseases, the main aim of classification is a communicational one. Before discussing the literature on "depression", it is thus important that we define what is implied by that term.

Communication is hindered if people rate depression differently. Variability between raters is dominated by three issues (Arbabzadeh-Bouchez and Lépine, 2003):

- Differences in the type of information obtained (information variance). Structured interviews aim to standardise the information elicited.
- Differences in the interpretation of information in terms of pathologies (interpretation variance). Explicit definitions of psychopathological terms aim to standardise the interpretation of psychopathological signs and symptoms.
- Differences in the definition of mental disorders (criterium variance). Reduction via standardised diagnostic criteria.

The second and the third point are addressed by international, standardised definitions of psychopathological terms and criteria for diagnoses. At present, these are the fourth edition of the Diagnostic and Statistical Manual (DSM IV, American Psychiatric Association 1994) and the tenth edition of the International Classification of Diseases (ICD 10, World Health Organization 1990).

DSM IV provides a list of features and states explicitly which and how many have to be present for a diagnosis of the various subtypes of depression. Here, we will concentrate on major depressive disorder (MDD), for which either a depressed mood, or anhedonia (a diminished interest or pleasure in almost all activities) are core diagnostic requirements. In addition, some of the following symptoms have to be present: vegetative (insomnia/hypersomnia, fatigue, weight gain/loss, psychomotor agitation/retardation); cognitive (perceived impairments in concentration); feelings of worthlessness or guilt; suicidal ideation. We will however at times also mention the two major, and despite their age still highly controversial, subclassification into endogenous/melancholic versus reactive/agitated depression (Nelson and Charney, 1981). While the former is dominated by features of retardation (Parker and Hadzi-Pavlovic, 1996), the latter tends to be associated with anxious phenomena. Both ICD-10 and DSM-IV are originally based on the Research Diagnostic Criteria (Spitzer et al. 1978, themselves based on the Feighner criteria; Feighner et al. 1972) and were designed specifically to decrease inter-rater variability. They were not supposed to be a definite classification, but rather a representation of the current knowledge. More recent definitions can claim additional accuracy with respect to patient populations as they incorporate findings from large, epidemiological studies (such as the Mood Disorders Field Trial for DSM IV; Keller et al. 1984).

The main tools apart from the categorical ones are "dimensional" and aim to assess both the presence and severity of particular features. They come in two forms. The first kind are self-report questionnaires, such as the Beck Depression Inventory (BDI, Beck et al. 1996), while others, such as the Hamilton Rating Scale for Depression (HamD, Hamilton 1960) are administered by an interviewer and also aim to assess which depressive features are present and how strong they are.

A number of major limitations in both the dimensional and categorical approaches are evident: the only data included are observations by raters, or subjective reports by individuals. Comparing it to other branches of medicine, this means that diagnosis is based on history (albeit structured and highly refined) and inspection. While these are very important medical tools, they very often are insufficient for an unequivocal choice amongst differential diagnoses. In fact, comorbidity between categories is immense (Blazer et al., 1994; Kaufman and Charney, 2000; Kendler et al., 2003b), arguing either against the existence of separate entities (Clark and Watson, 1991; Van Os et al., 1999) or, alternatively, that the categorical definitions have not solved the issue of reliability (Kutchins and Kirk, 1997; Bentall, 2003). Indeed, the classification 'Other / Unspecified depressive disorder' present in both ICD 10 and DSM IV is all but rare (Bentall, 2003). The categorical definitions and the diagnostic features are based on information mainly gathered in western countries, and there are important cultural differences in the reporting of symptoms that may again obscure the underlying pathology, even within western cultures, as they may assess what is a cultural phenomenon rather than a neurobiological disease process (Pilgrim and Bentall, 1999; Kleinman, 2004) — an interpretation further

supported by the fact that the categorical definitions have very little impact on choice amongst (biological) treatments Spitzer (1998); Parker and Manicavasagar (2005).

In our view, many of these issues are closely related to the fact that, in terms of the three approaches we described in section 1.1, only the cognitive one comes to bear in diagnosis. By this we mean that diagnosis rests to a large extent on introspective reports of symptoms by the affected person (although weight is given to signs, which are observer-rated). However, insights from biological and behavioural research have yet to be incorporated in diagnosis — in fact, they have to a large extent been validated by their concordance with the cognitively-derived categorical nosology. Up to a point this is desirable — psychiatry, just as any other branch of medicine, should be guided by what the patient or its entourage report as being disturbing. However, an essentially subjective definition is likely to obscure the view of underlying disease processes because we have extremely little understanding of the very complex processes involved in consciousness. It is therefore important to validate insights from the biological and behavioural approaches in their own right, by their covariance with cognitive measures during the progression of the disease, but also by their own sensitivity to treatment and aggravating factors; and by their ability to provide a stable sub-classification. Fortunately, this has been recognised in the literature and there have been important efforts in finding alternative "endophenotypes" (Hasler et al., 2004), so far centred on clinical signs (Parker et al., 1994), imaging (Mayberg, 1997), neuropsychology (Chamberlain and Sahakian, 2005) and genetics (Kendler et al., 2003b; Caspi and Moffitt, 2006). While the results from such studies have been very supportive, they have, so far remained within their theoretical framework and not yet attempted to integrate knowledge from the three approaches.

Here, we are motivated by the fact that a characterisation of depression within the framework of affective decision making would only integrate the various approaches, but might also form the basis for a nosological system that might be helpful in improving some of the above issues. A nosology based on performance of tasks that probe the various decision making systems could conceivably reduce the information variance by complementing interviews with computerised neuropsychological and psychophysical tests (akin to the CANTAB test battery; Robbins et al. 1994) and it might reduce the interpretation variance because it is defined normatively, in terms of normal brain function. We will see that there is evidence that all the systems mentioned in section 1.2 are involved in depression. It is tantalizing to think that sub-classes of depression might map onto particular decision-making systems, or that involvements of particular systems would increase sensitivity to particular treatments — say involvement of the goal-directed system to psychological therapy, motivational deficits to dopaminergic manipulations and issues in the Pavlovian systems to selective serotonin reuptake inhibitors (SSRIs).

As a very first, emphatically small, step in this direction, we will thus concentrate on those aspects of depression which are most immediately interpretable in a reinforcement learning framework — anhedonia, and feelings of worthlessness or helplessness — and neglect most of the vegetative and purely conscious symptoms. We will, however take a very broad approach to these three and describe their various facets in terms of each of the systems of section 1.2. Reinforcement learning provides a description of behaviour, rather than conscious reports. Therefore, it applies equally to animal and human behaviour, and the results of this thesis will rely extensively on the more thorough experimental designs possible in the animal

work (see section 2.9.1 for a discussion of the validity issues of animal models in the present context).

2.1.3 TREATMENT

For the present purpose, our main interest in treatment is the view it gives us on the mechanisms underlying depression. The two main types of treatment are pharmacotherapy and psychotherapy. We will review evidence from pharmacotherapy in section 2.1.4, and evidence from psychotherapy mainly in section 2.4. Pharmacological, psychotherapeutic and electroconvulsive therapy together are effective in a large fraction of patients, the success rate of pharmacotherapy and psychotherapy alone being consistently around 60%. Psychotherapy has not been shown to be effective in severe depression, but appears to be as effective as antidepressants and better than placebo in milder forms of depression, especially in primary care settings where patients tend to suffer from minor forms of depression (Gloaguen et al., 1998; Mulrow et al., 2000; Simpson et al., 2003; Hale, 2005). For major depression, antidepressants are the first-line treatment, and combined therapy is mainly considered in severe cases that do not respond to pharmacotherapy.

2.1.4 NEUROMODULATORS

Neuromodulators are heavily implicated in emotions. They also have an ancient affiliation with affective diseases. In humans, the data come in three types: assessments of neuromodulator function in people with depression, pharmacological effects on depressed people, and pharmacological effects on normals or recovered depressed people.

Asked to vote for the most likely neuromodulator irregularity in depression, based on the psychological findings and on facts like the high incidence of depression in PD (approximately 50-90%; Kapur and Mann 1992; Mentis and Delalot 2005), many have been compelled to choose dopamine (Randrup et al., 1975; Willner, 1985b; Depue and Iacono, 1989; Heinz, 1999; Naranjo et al., 2001), while others, assigning more weight to pharmacotherapeutic advances, might have chosen serotonin or noradrenaline (Lapin and Oxenkrug, 1969; Mann, 1999; Nutt, 2006). However, it has become clear that disturbances of tonic levels of a single neuromodulator cannot account for the data. Rather, we must think about the interplay between the various systems, and how they can jointly account for the many facets of depression. We will therefore review evidence on neuromodulators in the sections to which they have been most directly linked.

2.2 PRIMARY REINFORCER SENSITIVITY

Before discussing information relevant to particular affective systems, we need to review some evidence on the primary sensitivity to reinforcers themselves, as this will have implications for the interpretation of evidence relating to the systems. The primary sensitivity to a reinforcer is the strength of the reinforcer. For example, analgesia alters the primary sensitivity to painful stimuli. These are effects that precede any effects due to decision-making systems.

2.2.1 REWARD

Anhedonia is one of the core and most specific symptoms of depression (Costello, 1972; Willner, 1985b, 2002; Willner et al., 1987; Cloninger, 1987; Clark and Watson, 1991; Hasler et al., 2004). It is defined as an inability to enjoy things in life, for example: "I have a sort of uncanny feeling. I know what I am reading is amusing but I am not at all amused by it." (Sims, 2003). Anhedonia is assessed by all major dimensional tools, on which most therapeutic evidence is based and a large fraction of the animal modelling work that concentrates on face validity and construct validity is built around animal equivalents of anhedonia (Frazer and Morilak, 2005; Willner, 1997). Both the categorical and the dimensional tools use a verbal, conscious and subjective assessment by the affected individual to measure anhedonia.

The overall question pursued in the literature review is whether such subjective statements are indeed indicative of changes to the brain's affective decision making systems, and if so, what the nature of these changes is. The affective decision making framework outlined in section 1.2 is behavioural, and as such we are in search of a behavioural analogue to anhedonia — an effort which dates back to research in the seventies which already suggested that anhedonia may be due to a decreased sensitivity to rewards (see Costello 1972; Akiskal and McKinney 1973, 1975; Blaney 1977; Lewinsohn et al. 1979 for early reviews). More precisely, there are two definitions relating to rewards that are directly relevant to this endeavour. The first is its functions as a *reinforcer*. It is a strictly behavioural definition which defines as reward any outcome which will increase the future probability of the very action that led to it. As a caveat we have to keep in mind that this is a useful definition in a laboratory setting, but it is unclear how useful it is in a more general setting. Secondly, *motivational states* relate to the prediction of rewards (reinforcers). In states of high motivation, organisms predict that actions are available which lead to rewards, and thus that on average much reward can be obtained. There are two more concepts that are not stated in behavioural terms. One is that of hedonic impact. It is the conscious appraisal of the positive aspect of rewards. Finally, it may be that anhedonia really is a conscious assessment of the amount of reward achieved (predicted) by an individual — maybe compared to an aspired level of rewards. Of course, these are all tightly linked (Hull, 1943; Blaney, 1977; Lewinsohn et al., 1979; Willner, 1985b; Berridge and Robinson, 1998; Niv et al., 2007).

This section is organised as follows: firstly, we will review evidence for a decreased sensitivity to rewards in depression. Importantly, the concept of anhedonia has often been coupled with an asymmetry: the effectiveness of rewards is decreased, but that of punishments either unchanged or increased. Much of this research is based on memory biases, which we will only review very briefly. The second section will review evidence on sensitivity to punishments.

2.2.1.1 EXPLICIT MEASURES

Naturalistic population studies in which people fill out frequent questionnaires during their daily life provide good evidence that depressed people report and remember fewer rewards (reviewed by Lewinsohn et al. 1979 and Layne 1980) — as might be expected from the inclusion of anhedonia as a core diagnostic symptom. Importantly, these studies find a dissociation

between rewards and punishments compared to healthy controls: rewards seem to have less effects, punishments more, and such effects are not found for psychiatric controls (though see Williams 1992). More recent studies, employing improved methods, have essentially replicated these findings (e.g. Myin-Germeys et al. 2003; Jim Van Os, pers. comm.). An asymmetric skew towards lessened perception of rewards but not punishments has also been found in laboratory settings (Wener and Rehm, 1975; Buchwald, 1977; Nelson and Craighead, 1977): Subjects asked to rate the amount of positive feedback received report less the higher their (dimensional) measure of depression.

Indeed, a negative bias in tests of memory is one of the most replicated findings in depression. While it is present in both implicit and explicit tests, mood-specific memory biases, are not specific to depression. There is very good evidence that mood by itself biases mnemonic processing of valenced items, whether in depression or not (Blaney, 1986). Furthermore, for a thorough understanding of these effects (and also of their putative causative role in depression), a better understanding of the relationship between reinforcement learning, moods and memory is necessary. It may be that mnemonic effects are most important in the maintenance of depression (see for example theories centred on rumination; Nolen-Hoeksema 1991).

2.2.1.2 IMPLICIT MEASURES

The above findings have nothing to say about the locus of dysfunction associated with the expression of anhedonia: is it the reward process, or the interaction between the reward and the mnemonic systems? Early work tried to ascertain this again through verbal reports (DeMonbreun and Craighead, 1977), but more recently a number of approaches that are not obscured by reports have been used. The findings are rather contradictory. Studies that compare patients (with a DSM-III or DSM-IV diagnosis of depression) to healthy controls find that they show symmetrically lessened facial electromyographical responses (Greden et al., 1986), evoked response potentials (Deldin et al., 2001) and galvanic skin responses (Rottenberg et al., 2002) to stimuli of both positive and negative valences, though these studies did not demonstrate a correlation with a dimensional measure. Others measure the startle response during the viewing of affective imagery and find a full reversal of the effect of positive and negative emotional stimuli (Allen et al., 1999): while mildly depressed and non-depressed (BDI$<$ 29) subjects show an increasing startle response for more negatively valenced stimuli, the opposite is true for severely depressed subjects (indicating an insensitivity to negative, and a high sensitivity to positive imagery). Notably, this occurs despite similar explicit ratings of the imagery by all subjects. Finally, there is also evidence for an asymmetric effect on positively valenced as opposed to negatively valenced words (go/nogo reaction time to happy words were increased in depressed subjects, but reaction times to sad words were unaffected; Murphy et al. 1999; Elliott et al. 2002; Erickson et al. 2005).

Thus, the explicit reports by themselves are marred by reporting biases and thus can tell us nothing about any underlying changes to reward systems. Even the evidence on implicit measures of memory function is complex, as it is unclear whether the effects occur at the storage or at the retrieval time. The evidence on the implicit measures concurs in the finding that responses to rewards seem impaired, with the possible exception of the study by Allen et al.

(1999). However, it is unclear whether the effect symmetrically also affects punishments. We will return to this issue below, where we will argue that overall the evidence leans towards a symmetric impairment.

2.2.2 PUNISHMENT

Depression is not just not fun. It is a highly aversive state. We saw how it is one of the main actors in suicide. Sad mood, many would argue, is normal, even healthy, when life takes nasty turns, and maybe depression when life turns horrid (Nesse, 2000).

Here, we first review evidence that depression is associated with stress. We then turn to the effect of particular stressors in depression. A number of studies looked at autonomic responses to aversive and negative (but not painful) stimuli. As those known to us also include positive stimuli, we reviewed them in section 2.2.1. Those studies that eschewed conscious reports, in contrast to those that did not, generally agreed upon finding a decreased sensitivity to negative events.

2.2.3 STRESS: CORTISOL LEVELS

Stress is the main known aetiological factor in depression. There is good evidence that negative life events frequently precede depression (Bandelow 2003; further reviewed in section 2.8) and that depressed people express more distress upon experience of even minor stressors compared to healthy and psychiatric controls (Lewinsohn et al., 1979; Myin-Germeys et al., 2003). A large body of work has concentrated on cortisol, the main hormone coordinating the body's stress responses. If the state of depression is related to a disorganised, potentially overly strong stress response (Selye, 1984), then surely cortisol, the body's main stress hormone, should provide a good measure of this. Indeed, 50-60% of depressed persons have increases in baseline levels of adrenocorticotrophic hormone (ACTH), the main releasing hormone of cortisol, and also cortisol itself, probably due to a hypersensitivity of the adrenal glands. ACTH is in turn released by corticotropin-releasing hormone (CRH), both of which are under negative feedback regulation of cortisol, and it is the negative feedback which appears dysregulated in some depressive persons: When dexamethasone, a synthetic glucocorticoid, is given prior to CRH, control subjects respond with very little changes in ACTH, but depressed patients respond with a large surge of ACTH (Bandelow, 2003), arguing for an impaired negative feedback at the level of the pituitary (though see Gold et al. 2002). On the other hand, CRH on its own produces an ACTH response in controls, but not in depressed subjects. Both of the anomalies disappear with remission (Ströhle and Holsboer, 2003).

Studies in primates and rodents support the notion that a dysregulation of the stress response compounds the effects of stressors (Koolhaas et al. 1999; Sapolsky 2004, 2005; Korte et al. 2005, but see also Lechin et al. 1996), but a question of major interest has been whether these changes *cause* depression, or are themselves a *consequence* of depression. Strickland et al. (2002) looked at cortisol responses of nearly 500 women in a community setting. They assessed the levels of current stress, and the history of life events, and measured morning and evening salivary cortisol. They found that the cortisol levels of women with depression did not dif-

fer from that of control subjects, and that their cortisol responses to recent severe life events were equal. However, they also found that cortisol levels in those women with depression and chronic stress were increased, whereas chronic stress had no effect on the cortisol levels of control subjects. Thus, it seems that a high cortisol response to a severe stress is not associated with depression in the community, but that experiencing chronic stress when depressed leads to an elevated cortisol response — depression here appears to cause or permit the cortisol changes, rather than the opposite.

2.2.3.1 Pain

Pain, a very immediate form of punishment, is also associated with depression — indeed, in the 18th and much of the 19th century hypochondria replaced melancholia as the term of choice for what may now be termed depression. Because it is inducible in a controlled, ethical manner, it is the most accurate measurement of sensitivity to primary punishments available in humans and its association with depression further lends it credibility.

Pain and human depression have a complex and only very partially explored relationship. *Prima facie*, pain seems related to the induction, or at least the initiation of depression: it is estimated that 80% of patients with MDD initially present with pain of some sort (Naber, 1988); chronic pain is a frequent precursor to depression (Lautenbacher et al., 1994); pain is reported by most people suffering from depression (> 50% Dworkin et al. 1995) and improves with antidepressant medication (Katona et al., 2005). However, the bond between pain and depression is in no way exclusive — similar increases in pain thresholds are present in a number of psychiatric conditions (e.g. mania, anorexia nervosa and bulimia nervosa, but not panic disorder (Lautenbacher et al., 1994, 1999)). Neuroticism, which has a long-standing relation to both anxiety, pain and the "reactive" conceptions of depression (Eysenck, 1997) also affects the perception of pain (Vossen et al. (2006) and references therein).

It is also well-known that during depression, pain thresholds are found to be increased. In an early report, Hall and Stride (1954) described increased heat pain thresholds. They tested neurotic and depressed patients (no further details about patient selection). They found that thresholds for judging the stimulus as painful are enhanced in people diagnosed with endogenous depression, and that they decrease again after response to ECT treatment. As a control task, they also ask subjects to judge whether less potent stimuli are just noticeable or distinctly noticeable, and find no difference between depressed and normal subjects. More recent studies have replicated the effect and confirmed the specific increase in the pain threshold with the use of more sound methodology and analysis tools (signal detection analysis, contact heat rather than electric shock, non-noxious controls Davis et al. 1979; Lautenbacher et al. 1994; Buchsbaum 1979, though see also Dworkin et al. 1995, who re-analyse data from the 70s). However, these findings do not correlate with dimensional measures (only Lautenbacher et al. 1999 report that it is less prominent in patients with pain complaints) and are not affected by naloxone (i.e. are not mediated by central opioids; Lautenbacher et al. 1994). Against these positive findings, there are also reports that the changes in pain sensitivity are specific to certain modalities (Bär et al., 2005) and may not apply to life events in general (Lewinsohn et al. 1979 and references therein).

Overall, thus, we conclude that there is strong evidence for *decreased* sensitivity to primary punishments in depression, although there is most likely an enhanced stress response which parallels the enhanced conscious appraisals of stress and punishments.

2.3 PAVLOVIAN ACTIONS

There is no direct information in human depression on the Pavlovian decision making and action selection system, but there are indirect indicators. Firstly, some assessments of primary reinforcer sensitivity involve reflexive actions which are presumably under Pavlovian control. For example, GSR responses might be seen as internally directed Pavlovian actions. Decreased GSR responses (Rottenberg et al., 2002) could then indicate either decreased primary sensitivity, or a decreased coupling between actions and the kinds of reflexive, preparatory actions the Pavlovian system is supposed to control. Secondly, some response biases might be interpreted as Pavlovian biases. For example, depressed subjects respond more rapidly to negatively than positive valenced words in a go/nogo task, and this difference is not present in healthy humans (Murphy et al., 1999) or reversed (Erickson et al., 2005).

However, the main insight into Pavlovian systems in depression comes from an interpretation of two of serotonin's putative roles in normal behaviour — inhibition of actions and reporting of negative values. In chapter 5, we will explain how the interaction of these two roles might explain why *de*creases in serotonin levels might result in states akin to depression. We therefore review the evidence of serotonin involvement in depression here.

2.3.1 SEROTONIN IN DEPRESSION

Prima facie, the fact that serotonin is increased by stress (Takase et al., 2004) and inhibits actions (Soubrié, 1986; Gray, 1991) might lead to the hypothesis that 5HT levels in depression should be *in*creased. This is a conclusion one might also draw from the association of the less efficacious 5HT transporter allele with depression (Caspi et al., 2003). However, the opposite appears to be true: There is evidence for lowered levels of 5HT in people currently experiencing depression (this is the original indoleamine hypothesis; Lapin and Oxenkrug 1969) and lowering 5HT leads to a re-experience of depressive symptoms (Delgado et al., 1994; Smith et al., 1999). Antidepressants inhibit the 5HT reuptake mechanism, thereby increasing levels of 5HT (Maudhuit et al., 1997; Millan, 2006), and appear to be more efficient in carriers of the (more effective) *l/l* 5HT transporter polymorphism (Whale et al., 2000; Lotrich et al., 2001; Yu et al., 2002).

Direct measurements in cerebrospinal fluids (CSF) have provided only scant evidence. Studies looking at levels of serotonin's major metabolite, 5-HIAA (5-hydroxy-indole-acetic acid) have shown in equal numbers heightened and lowered levels when compared to controls, and these levels have not correlated with severity of depression, or with psychomotor symptoms (Willner, 1985b; Mann, 1999; Sibille and Lewis, 2006). Somewhat more consistent evidence for a decrease in 5HT activity comes from prolactin responses to fenfluramine challenge[1], which

[1] Fenfluramine releases 5HT from stores and inhibits the reuptake. This 5HT in turn leads to release of prolactin.

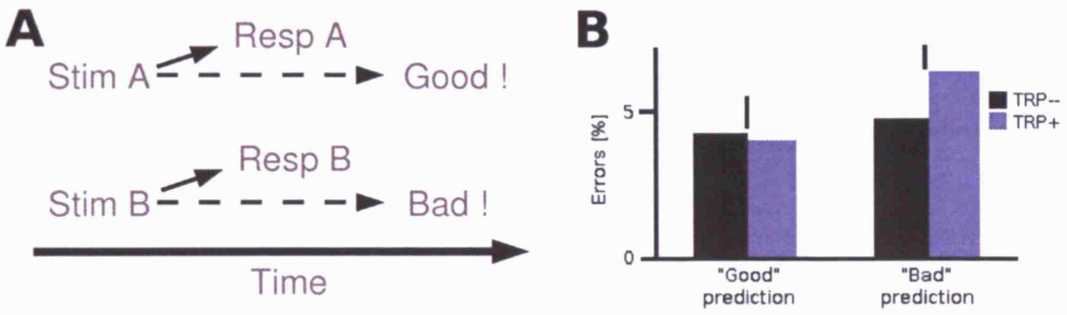


FIGURE 2.1: Pavlovian inhibition mediated by serotonin. **A**: Experimental setup. Two stimuli were presented. Subjects were instructed always to produce response A for stimulus A and response B for stimulus B. They were not given any feedback. Stimulus A was followed, regardless of whether the correct response was emitted, by an appetitive outcome, while stimulus B was followed by an aversive outcome. The outcomes thus do not reinforce the correct responses. **B**: Control subjects performed more errors on stimulus B, which was follwed by an aversive outcome. This excess was abolished by tryptophan depletion. Adapted from Robinson et al. (2007).

are blunted in both depressed subjects and subjects with a history of depression alike (Mann, 1999). Some have suggested a bimodal distribution, maybe indicating features present only in a subpopulation of depressed people (Korf and van Praag, 1971a). Indeed, the strongest correlation is with lack of inhibition, suicide or suicide attempts and with aggression (Träskman et al., 1981; Willner, 1985b; Deakin, 2003).

The strongest evidence in favour of the involvement of low tonic 5HT levles in depression are the efficacy of selective serotonin reuptake inhibitors (SSRI) in treating depression (Rang et al. 2000; Millan 2006; Sibille and Lewis 2006; Hariri and Holmes 2006; though see also the serotonin reuptake enhancer tianeptine Sarek 2006) and direct manipulations by tryptophan depletion (TrD). TrD involves the administration of a drink that contains all amino acids but tryptophan and reliably induces a strong depletion in plasma and brain tryptophan and serotonin levels. While TrD does not reliably induce depression in all subjects, it does induce depressive symptoms in subjects with predisposing factors such as a family history of depression; the short 5HT transporter allele (short allele of the 5HT transporter locus polymorphic region; 5HTTLPR) (Young et al., 1985; Bell et al., 2001; Neumeister et al., 2002); or a history of depression (Delgado et al., 1990; Smith et al., 1997; Moreno et al., 1999). These effects are mainly due to reversals of the antidepressant treatment itself (Delgado, 2000; Iversen, 2005): 70% of patients who have responded well to SSRIs suffer a relapse after tryptophan depletion (Delgado et al., 1990), whereas only 20% of those who responded well to selective noradrenaline reuptake inhibitors (SNRI) experience a relapse (Delgado et al., 1994). α-methyl-para-tyrosine (AMT) inhibits the first step in catecholamine synthesis, and leads to decreases in brain NA akin to the depletion in 5HT seen after tryptophan depletion. AMT re-induces depression in 81% of patients that had responded well to SNRI, but only 19% of patients that had responded well to SSRI (Miller et al., 1996a,b). Thus, NA depletion reverses SNRI treatment effects and

TrD reverses SSRI effects. Both reverse effects of mixed drugs (Delgado, 2000; Iversen, 2005). These findings argue strongly that low 5HT levels do not induce depression by themselves, and increases in 5HT are not necessary for remission. Nevertheless, some people are sensitive to decreases in 5HT, and these are the ones who respond to pharmacological inhibition of the 5HT transporter (5HTT) by SSRIs.

In people without predisposing factors, TrD does not elevate standard measurements of depression. But it does reproduce important, subtler, aspects of the depressive status. For example, TrD has effects related to value that are reminiscent of effects seen in depression: it increases reaction times to happy but not sad words in a go/nogo task (Murphy et al., 2002); induces a negative memory bias (Klaassen et al., 2002); abolishes reward-induced reaction time speeding (though only in carriers of the short 5HTTLPR allele Roiser et al. 2006) and increases negative mood responses to uncontrollable stress (Richell et al., 2005). It also has effects akin to those seen in animals, speeding up temporal discounting (Doya, 2000; Schweighofer et al., 2006, 2007) and impairing reversal learning (Rogers et al., 2003). Finally, there is one report consistent with the interpretation that TrD specifically affects actions based on aversive Pavlovian values (figure 2.1; Robinson et al. 2007). As usual, there are also findings that are hard to interpret. Cools et al. (2005) report that impulsive subjects initially respond more rapidly to rewarded events than less impulsive subjects, but that after TrD the opposite is true: impulsivity is now anticorrelated with the effect of rewards on reaction time (see also Dalley et al. 2002). We report this study here mainly because the effect of TrD is very strong. And finally there are findings that are outright contradictory: TrD reduces people's ratings of amphetamine-induced "highs" (Aronson et al., 1995) whereas it is increased in depression (Tremblay et al., 2002).

Data on 5HTT levels — the main target of the majority of antidepressants — are relatively consistent. A large number of studies have shown that platelet 5HTT levels are consistently low in depression (Ellis and Salmond, 1994). Using more direct measurements, post-mortem analyses (Mann et al., 2000), single-photon emission computed tomography (SPECT; Malison et al. 1998) and positron emission tomography (PET; Parsey et al. 2006b), decreased 5HTT levels (the binding potential is reduced by 20%) in brainstems of depressed people are found. However, neither of these latter two measures correlates with any dimensional measure of depression. Interestingly, 5HTT levels in neither study correlated with suicide attempts or were affected by a history of antidepressant medication. Interestingly, the levels of brainstem 5HTT are not affected by 5HTTLPR status (Parsey et al., 2006a).

We will see that our treatment of the interaction of the various roles of serotonin do not necessarily ascribe it a more prominent role in depression as compared to anxiety. Indeed, it might well be that 5HT is specifically associated with the comorbidity of depression and anxiety. After all, SSRIs are first-line treatments not only for depression, but also for most anxiety disorders. In support of this, 5-HIAA levels correlate well with features of anxiety in depression (Willner, 1985b); TrD can increase measures of both anxiety and depression (Smith et al., 1997); L-tryptophan challenges yield blunted prolactin responses in non-melancholic, but not in melancholic patients (Price et al., 1991); the short allele version of the 5HTT is directly associated with anxiety rather than depression (Lesch et al., 1996); and functional brain measures associated with the short 5HTTLPR allele correlate strongly with temperamental indices of anxiety (Pezawas et al., 2005). However, no relapse of anxiety analogous to that of depression has

as yet been reported with TrD, although there are effects on experimental panic (Klaassen et al., 1998; Schruers et al., 2000; Miller et al., 2000). The effects of AMT and tyrosine depletion on anxiety have, to our knowledge, not been explored.

2.4 GOAL-DIRECTED DECISIONS AND CONTROL

Cognitive theories of depression have at their core the notion that the central dysfunction rests on judgements that reinforcers are *unlikely* to be earned or observed, rather than that they are negligibly large. The former is at the heart of what defines control. Control is present in situations in which desired outcomes are judged to be highly likely if an appropriate action is chosen. Absence of control means that the desired outcome cannot be achieved, either because the outcome is unlikely for all possible actions, or because the outcome is unlikely under the particular actions available. For 40 years ideas relating depression to a perception of no control have been very prominent and resulted in research far too voluminous to fit into a thesis, so we will here only give a very superficial overview of the main issues of direct relevance to reinforcement learning.

2.4.1 LEARNED HELPLESSNESS

The main ideas came in two guises, which have now come to be seen as closely related: One is Beck's cognitive theory of depression, the other the theory of learned helplessness and its successors. Seligman and Maier's learned helplessness (LH) theory (and, nearly simultaneously, Jay Weiss' behavioural depression theory; Weiss et al. 1980, 1981) arose from animal behaviour experiments, but quickly motivated experiments in humans (Overmier and Seligman, 1967; Seligman and Maier, 1967; Miller and Seligman, 1975; Maier and Seligman, 1976). The basic finding (see figure 2.2) involves three rats and two experiments on subsequent days. On day one, the "master" rats are exposed to a series of shocks that are unpredictable, but which the master rats can switch off, for example by turning a wheel (escapable shocks, ES). The "yoked" rats are exposed to precisely the same shocks as the master rats, i.e. their shocks begin and end at the same time. The only difference is the action-outcome contingency: yoked rats do not have control over when their shock is switched off (inescapable shocks IS). On day two, both groups of rats are given escape training: Electric shocks are again delivered at random times, but both groups can terminate them by escaping to the other partition of the shuttle box (i.e. now both groups have full control). Master rats learn this readily (they do not differ from a third group of rats that was not exposed to any shocks) whereas yoked rats fail to do so. Seligman and colleagues argued that the crucial variable is control: a perception of no control (instilled by exposure to severe, inescapable shocks) results in a failure to use (negative) reinforcements to guide actions (they see this as cognitive deficit), presumably because actions are not expected to prevent the punishing outcome. We will see in chapters 3 and 4 one of the most crucial aspects of the data is that even exposure to uncontrollable *rewards* can induce escape deficits (Goodkin, 1976; Overmier et al., 1980).

The authors went on to argue that a perception of no control would decrease a person's general motivation and, crucially, that it would result in depressed or anxious mood. This strong

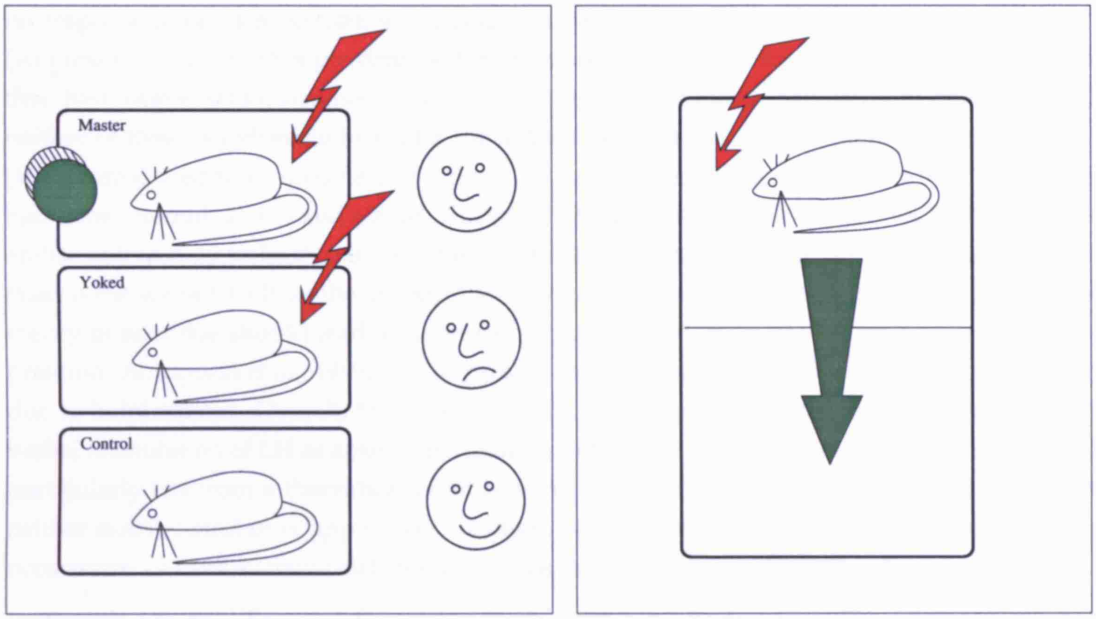


FIGURE 2.2: The learned helplessness paradigm. Three sets of rats are used in a sequence of two tasks. In the first task, rats are exposed to escapable or inescapable shocks. Shocks come on at random times. The master rat is given escapable shocks: it can switch off the shock by performing an action, usually turning a wheel mounted in front of it. The yoked rat is exposed to precisely the same shocks as the master rat, i.e. its shocks are terminated when the master rat terminates the shock. Thus its shocks are inescapable, there is nothing it can do itself to terminate them. It may have a wheel mounted in front of it, but this wheel cannot be turned. A third set of rats is not exposed to shocks. Then, all three sets of rats are exposed to a shuttlebox escape task. Shocks again come on at random times, and rats have to shuttle to the other side of the box to terminate the shock. Only yoked rats fail to acquire the escape response.

link to depression was supported by early human analogues of the animal LH experiments which often (Miller and Seligman, 1975; Roth and Kubal, 1975), though not always (Willis and Blaney, 1978), claimed to induce states that lead to increases in e.g. BDI scores. However, in order to induce helplessness, subjects had to be told explicitly that they did badly at one task, before they came to perform poorly on another, and it has been argued that this is an unavoidable confound, given that items on these scores explicitly assess expectations of performance (Blaney, 1977). Research into the effects of exposure to controllability and into the correlations between helpless behaviours and measures of mood like the BDI also produced conflicting findings (Blaney, 1977). LH was also rapidly criticised for predicting patterns of symptoms that were not observed, such as inactivity and passivity in reactive depression (Huesmann, 1978; Buchwald et al., 1978).

The attributional reformulation of LH (Abramson et al., 1978) and its successors aimed to address the shortcomings. Fundamentally, these argue that "an expectation that highly desired outcomes are unlikely to occur or that highly aversive outcomes are likely to occur and that

no response in one's repertoire will change the likelihoods of occurrence of these outcomes" (Abramson et al., 1988) is proximal sufficient cause for depression. There are two aspects to this: first, unpleasant outcomes of actions are judged more likely than pleasant ones. Second, neither of these is judged to be under the individual's control. Furthermore, Abramson et al. (1978) introduced three qualifiers of control: for helplessness to ensue, perception of no control had to be internal, global and stable. This makes intuitive sense: the controllability has to generalise between all tasks the subject attempts to do. The only constant variable across all these tasks is the subject itself, so the uncontrollability has to be associated with the subject. A precise theory of how this should lead to depression appeared a decade later when hopelessness depression (Abramson et al., 1989) was suggested as a separate subtype of depression, putatively due to helplessness. Overall, the reformulations rectify earlier internal inconsistencies in the verbal formulation of LH as applied to humans, but do not alter the substance of the argument, particularly not from a theoretical point of view: subjects are thought to judge that they can neither exert control over appetitive outcomes, as they cannot increase the likelihood of their occurrence, or over aversive outcomes, because they cannot avert them.

Beck's cognitive theory of depression (Beck, 1967, 1987; Beck et al., 1979) arose from clinical work with depressed patients, from detailed descriptions of their thought processes, argumentations and emotional responses. The theory's core statements are that depressed people experience negative automatic thoughts and make systematic logical errors, overall resulting in long-lasting, deep-rooted negative attitudes termed "depressogenic schemas". The logical errors described are often about adherence to a negative belief about their own person in the face of evidence which on balance would lead others (the therapist) to hold (more) positive beliefs. This is a notion very closely related to that of a global, internal and stable negative attribution defined in helplessness and hopelessness theories.

There is now very good evidence for hopeless attributions: in a large meta-analysis, Sweeney et al. (1986) concludes that depressed people attribute negative events to themselves (they make internal, global and stable attributions), but attribute positive events to random outside factors (external, specific and unstable). In an early example, Rizley (1978) found that subjects with higher BDI scores, relative to subjects with lower BDI scores, thought effort and ability (internal causes) were more important when they failed, but less important when they succeeded (Lyon et al. 1999 shows the opposite for manic subjects). There are important relational issues, in that depressed subjects make these errors only for themselves, but not when judging other people's actions (Golin et al., 1977).

Furthermore, the success of psychotherapy, or "talking therapy" as it is often called, has provided strong validation for the cognitive theories on which it is based. Cognitive-behaviour therapy (CBT, Beck (1967); Beck et al. (1979); Padesky (1994)) aims to educate subjects into interpreting emotional evidence they experience in a way that puts them into a better light. However, none of these are bedded in the kind of specific terms that would be needed to gain insights into the underlying mechanism, nor have such specific issues been examined. However, all this data so far is confounded by issues of verbal report.

2.4.2 CONTINGENCY JUDGEMENTS

However, there is a somewhat different string of research on contingency judgement, or "depressive realism" which is more relevant (though note, it still relies on conscious judgement). Abramson et al. (1979) (see also Alloy and Tabachnik (1984); Alloy and Abramson (1988); Dickinson et al. (1984)) found that non-depressed subjects were more likely than depressed subjects to judge that there was a contingency between their action and some outcome when in fact there was none. There was no difference between the groups when there was either a positive or a negative contingency. Apart from complex effects of the precise experimental setup (see Msetfi et al. 2005; Wasserman et al. 1993 for a discussion of the dependence on inter-trial intervals), this is consistent with the notion that healthy subjects have a stronger prior expectation of action-outcome contingencies than depressed subjects, but it does not square up to the LH argument, which is about missed existent contingencies specifically involving reinforcing events, not overestimates by healthy controls of non-existing contingencies that do not involve reinforcers.

Furthermore, the verbal judgements are very rough assessments. To our knowledge, there is only one study which partially addresses the question in a behavioural context (the Must et al. (2006) study is a second example, but, short of building a full model, too complex to interpret). Murphy et al. (2001b) give manic, depressed and control subjects explicit information about the probability of two events, of which they have to choose one. They place bets in percent of their current total of points. If the chosen event occurs, this amount is added to their total score of points, otherwise it is subtracted. The authors find that, compared to controls, 1) manic but not depressed patients choose the less probable outcome more often; 2) both manic and depressed patients place higher bets in uncertain situations (risk seeking), but place lower bets in certain conditions (fail to exploit), resulting in less total earnings. This may indicate that manic patients assume too much control, but on the other hand it also argues that depressed patients do not differ from controls. It is unclear how to interpret the finding that both patient groups bet more than controls in situations of high uncertainty, but less in situations of low uncertainty.

2.4.3 PLANNING

Performance on the Tower of London task (Shallice, 1982) is an important measure of executive function and directly assesses the ability to search a tree for the best sequence of moves. Although this is a conscious planning task, it is not confounded by issues of verbalisation and report, and arguably reflects the underlying decision processes. Experienced chess players for example are more proficient at it than control subjects (Unterrainer et al., 2006): they solve more of the hard problems by spending more time deliberating and executing the sequence of moves more slowly (presumably still deliberating). Depressed patients are worse than controls (Beats et al., 1996; Elliott et al., 1996), and more so the harder the problem. They also spend increasingly more time thinking with harder problems, but without an ensuing increase in their capacity to solve the problems (see also Goodwin 1997). One possible interpretation of this is that they search in larger, more complex trees, which takes both longer and is harder. This may be because they search in subparts of the tree that control subjects rapidly dismiss (a process related to pruning (Baum and Smith, 1997)), or also because they explore more actions follow-

ing each move. We will show in chapter 4 that a prior expectation of no control could result in the latter.

Overall, thus, there is good evidence that people's reports about controllability are closely linked to the development and the remission of depression. To the extent that such judgements precede the onset of depression, they may have some causal role, although most patients still are missed by these measures. There is at present no strong direct evidence that a perception of no control actually implies any dysfunction of the goal-directed affective decision making systems we showed depended on explicit estimates of probabilities and action-outcome contingencies. However, the data reviewed renders this a distinct and interesting possibility.

2.5 HABITUAL LEARNING

So far we have seen convincing evidence that the depressed state can be associated with alterations in primary reinforcer sensitivity and reports about the extent of behavioural control. Here, we review evidence about habitual learning, or stimulus-response (SR) learning. Although habits have not been explicitly linked to depression (unlike to drug abuse or schizophrenia; Everitt and Robbins 2005; Smith et al. 2004), the immense richness of animal and human work has endowed it more than others to tests of reinforcement systems in depression that eschew conscious issues. We now ask whether there is a correlation between the conscious reports of anhedonia assessed by the various measures of depression and the ability to acquire appetitive or aversive habits. We will be particularly attentive as to whether the data discriminates between effects due to potential differences the primary reinforcer sensitivity and deficits specifically of acquisition. We will go through all studies known to us in some detail.

2.5.1 APPETITIVE HABITS

Most directly relevant information comes from the development of response biases in signal detection tasks, early on without, later with, an explicit reinforcement component. There are also a number of very early studies on conditioned reflexes in depressed subjects (Ivanov-Smolensky, 1925; Ban, 1964), but neither are the patient groups described sufficiently, nor do the results have enough specificity. Martin and Rees (1966) give control subjects and patients with endogenous or mixed depression a discrimination task, in which a light predicts a tone to which subjects have to respond as rapidly as possible. The more severe, endogenous, patients do not develop preparatory muscle activity even to stimuli that predict the occurrence of the reaction time stimulus perfectly. Reaction times are longest for endogenously depressed subjects, intermediate for subjects with mixed depression and shortest for controls. However, the authors acknowledge that patient selection and categorisation was complex and probably unreliable, and that slowness being a diagnostic sign (i.e. a feature assessed by the clinician) for endogenous depression is likely to confound the findings. Furthermore, the rewarding nature of correct responses, compared to the aversive nature of incorrect responses was neither controlled nor measured, acquisition curves are not presented, and the baseline motivation to take part in the task was not measured. Thus it is unfortunately unclear whether the differences between the groups reflect changes in primary reinforcer sensitivity (of either valence), habitual

response acquisition, the expression of habitual patterns, or also motivational issues. Of note is that here the behavioural patterns were somewhat dissociated from those reported verbally, arguing that behavioural and conscious measures may be differentially sensitive to the issues of interest here. Finally, Miller et al. (1975) were unable to replicate the effects in their own version of discrimination learning: depressed students differed in the verbal reports of their own performance, but not in their actual performance.

Henriques et al. (1994) had controls and depressed subjects perform a verbal memory task and give reinforcements. They include three different monetary payoff conditions: either subjects are initially given credit, and each error results in a subtraction from this credit, or they are not given initial credit, but earn for each correct word recognition. The third condition is a neutral control. Depressed patients only develop a response bias in the negative feedback condition, whereas normals only show it during the positive feedback condition. Thus they argue that depressed people can use punishment to "motivate" them, but fail to do so with reward. However, it is unclear whether their results really support their main claim, which rests on the fact that there is no difference between groups in the punishment condition, but that there is one between groups in the reward condition. The negative result seems to be due to a lack of power, as the effect size would appear relative large (figure 1 in Henriques et al. (1994)), which is acknowledged in their discussion. Their interpretation leaves unanswered the question of why normals do not alter their behaviour to an equal extent in the positive and negative conditions. The block design employed may have complex effects — maybe the effect observed is a side-effect of subjects having been told that they were given credit initially, and then being 'anxious' about losing it. It would therefore have been desirable to control for anxiety scores on this task. In the second paper, Henriques and Davidson (2000) report more conflicting results. They initially criticise the bias measure used in Henriques et al. (1994), but unfortunately do not reanalyse that data. They then go on to find results that differ in important ways. They find that depressed subjects (a clinical population) do not differ from normals in any of the payoff conditions individually, but that there is an interaction: The difference between the normal group and the depressed group is larger in the reward than in the punishment condition. Interestingly, there is a trend for the depressed group to show less bias in the reward and punishment conditions than in the neutral condition. Furthermore, their analysis also reveals that there is a strong effect of anxiety on the performance in the punishment condition. Unfortunately, rather than controlling for this through a hierarchical analysis, they exclude patients with a history of anxiety disorders, which given the strong comorbidity (not just historical, but also concurrent, see Ninan and Cummins (2003) and references therein) would not appear sufficient to exclude anxious processes as a cause.

The clearest results have come from Diego Pizzagalli's lab, who have also used signal detection tasks to tap into the reinforcement processes in depression. Pizzagalli et al. (2005) give subjects the signal-detection task depicted in figure 2.3. Subjects were not told when they committed errors, but on a fraction of correct trials were given positive feedback ("Correct! You have won 5 cents."). For one stimulus (the "rich" stimulus), positive feedback was given on 3/4 correct trials, for the other ("lean") on only 1/4 correct trials. In doubt, it thus became advantageous to assume the rich stimulus had been presented, and subjects did develop a robust response bias (figure 2.3B). Subjects with lower BDI scores developed an increasingly large

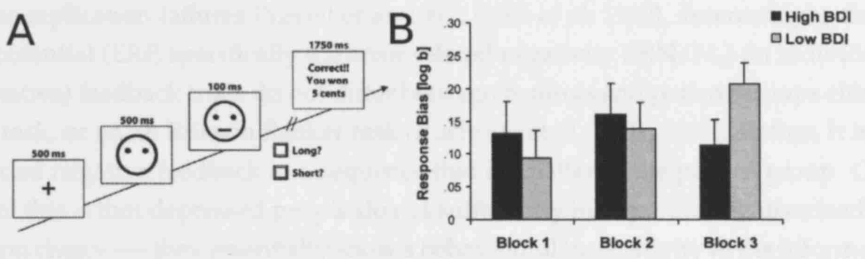


FIGURE 2.3: **A**: Asymmetrically reinforced discrimination task. After being presented with the stimulus, subjects had to indicate which mouth (long or short) had been presented by pressing the 'z' or '\' key. On a fraction of the trials they would be given positive feedback indicating how much money they had won (no negative feedback given). **B**: Response bias for subjects in the high BDI and low BDI group. The low BDI group develops a larger response bias towards the more rewarded response over the three blocks of 100 trials. Figure adapted from Pizzagalli et al. (2005).

response bias over the three blocks of 100 trials, and indeed the rate at which they did so correlated negatively with the BDI melancholic subscore (their figure 4). Thus, when uncertain, less depressed subjects were more willing to learn to go for the action which, on average, was associated with more reward. While this is very nice evidence that anhedonia is associated with some aspect of learning from rewards, the data as yet does not dissect the issue further: it is unclear whether the difference is due to differences in primary reward sensitivity, due changes in the learning process, due to uncertainty itself, or even due to some combination of these factors.

Overall, in conjunction with the evidence on measurements of responsiveness to rewards (section 2.2.1, it is at present more prudent to conclude that anhedonia is associated with decreased primary sensitivity to rewards, and that this itself accounts for the features of reward-related habitual acquisition seen in depressed patients. It may be possible to distinguish these issues with the employment of standard computational models of habitual learning (see appendix A), or by combining experiments on the acquisition of habits with those on primary reinforcer sensitivity. It would also be of great interest to replicate and understand the apparent asymmetry between rewards and punishments, both in healthy controls and in depressed people.

2.5.2 AVERSIVE HABITS

We are not aware of straightforward aversive learning studies such as fear conditioning or avoidance learning (though we will see some in the more complex context of learned helplessness below). However, there are a number of studies that include negative outcomes in an appetitive context.

It has been reported that depressed patients (DSM-III-R) are very likely to commit an error on the trial following negative feedback (in a variety of tasks). This is in comparison to both healthy and psychiatric controls (Beats et al. 1996; Elliott et al. 1996, 1997; Steffens et al. 2001;

but see also replication failures Purcell et al. 1997; Shah et al. 1999). Interestingly, the evoked response potential (ERP, specifically the error-related negativity ERN/N_e) on individual negative (or positive) feedback trials do not differ between controls and patient groups either on the go/no-go task, or on an Erikson flanker task (Ruchsow et al., 2004, 2005). Rather, it is the ERP on the second negative feedback in a sequence that is smaller in the patient group. One interpretation of this is that depressed people do not sufficiently incorporate negative feedback into future action choice — they essentially show a behavioural insensitivity to the information provided by negative feedback, and fail to avoid the punished behaviour. Somewhat unexpected are findings that carriers of the short 5HTTLPR allele (which is associated with depression) show larger ERNs Fallgatter et al. (2004), although this might relate to the allele's more direct relationship with anxiety.

Further evidence along this line comes from complex decision-making tasks that are known to rely heavily on prefrontal lobes. Must et al. (2006) report results on a modified version of the Iowa Gambling Task (IGT). In the IGT, subjects must make choices between decks of cards. One deck of cards always gives a large reward, the other a smaller reward. In the classical task, the first deck also gives rare large losses, and is overall disadvantageous, while the second gives more frequent but yet smaller losses. Depressed patients earn less rewards than control subjects on this task, i.e. fail to learn to avoid the risky, large-reward deck. This behaviour is famously associated with frontal lesions (Bechara et al., 2005). In the inverted version, one deck gives large punishments at each trial, but intermittently even larger rewards, and is advantageous overall, while the other deck gives smaller punishments on every trial, but also provides smaller intermittent rewards. Thus, to successfully choose the rich deck, subjects have to overcome the aversion to large losses. Depressed patients earned as many rewards as control subjects here, i.e. learn to chose the large reward / large punishment deck more often. The results are complex. They do not show that these effects correlate with any particular dimensional / severity measures. The task involves judgements about probabilities and sizes of two types of events, and explicit conflict. Also, there are issues with the task in general (it may have more to do with reversal learning than impulsivity, Fellows and Farah 2005) and with its interpretation (Maia and McClelland, 2004). Nevertheless, they are interesting: depressed subjects are not simply indifferent to either rewards or punishments (otherwise they would choose the deck associated with smaller punishments or larger rewards in both cases, or respond randomly throughout). They fail to avoid the deck with rare, large losses (they are either insensitive to the size of the losses, or underestimate their likelihood), but do not fail to choose the deck with rare large wins (they are sensitive to the reward size and do not underestimate its probability). Thus, a simple interpretation of the results is that depressed patients respond normally to rare rewards, but not to rare punishments.

The Wisconsin Card Sorting Task (WCST) is a complex task, in which subjects must first infer a correct set of actions, and then alter these as reward contingencies are changed. The initial step of inferring the correct action set is beyond our scope as it involves choosing actions in order to gain maximal information, rather than just choosing actions that maximise rewards. However, the set-shifting stage is of interest. Patients with dysphoria (Channon, 1996), melancholia (Austin et al., 1999) and a DSM-III-R diagnosis of major depression (Goldberg et al., 1993) make more perseverative errors than controls, i.e. they fail to use negative feedback to

alter their response strategy. Due to the complex nature of the task, this may both be because they are less behaviourally sensitive to the negative reinforcers, or because they are unable to choose an informative alternative action (in fact, the latter is more likely). A simpler version of a similar task is reversal learning. Courville et al. (2004) and Hampton et al. (2006) point out that it involves learning of a higher-order structure. Initially, subjects need only associate one stimulus with one action. As the reward contingencies are reversed, subjects update their strategy over a number of trials. Over time, subjects learn to reverse more rapidly: They effectively learn that the task is nonstationary and that two different response contingencies apply at different times. Reversal learning in marmoset monkeys and humans is sensitive to prefrontal serotonin depletion and lesions of the orbitofrontal cortex (Dias et al., 1996; Rogers et al., 1999; Cools et al., 2002; Clarke et al., 2004, 2005). We do not know of any studies relating performance on precisely this task to measures of depression, but the experiment by Murphy et al. (1999) reported has a related facet. Their emotional go/no-go task consisted of 8 blocks, in which sad or happy targets were alternatively designated as targets. Depressed patients took longer to adjust to the reversals, which were instructed and supported by negative feedback for errors (no positive feedback). There was no interaction with valence.

A rather different finding — of increased sensitivity to stress in subjects with high BDI scores — comes mainly from the signal detection tasks mentioned above (Henriques et al., 1994; Henriques and Davidson, 2000), and from verbal reports. However, we mentioned the issues with this data. A task by Bogdan and Pizzagalli (2006), still fundamentally appetitive, gives the clearest evidence for increased (but nevertheless normative) sensitivity to punishments: subjects are instructed to report which signal was presented, but are rewarded more often for correct answers on one than the other stimulus. When given overall performance feedback and told they do badly, the response bias decreases — as one would expect, for now the reinforcement schedule is more even and depends on both stimuli. However, the decrease in response bias (i.e. the sensitivity to negative feedback) correlated with BDI score, indicating an increased sensitivity to punishment. Very importantly, the effect did not correlate with MASQ (Watson et al., 1995) measures of general or anxious anxiety.

Thus, depressed people appear to be less willing to use negative feedback about specific actions to avoid those actions in the future, which is consistent with the results from the sections on primary punishment. The signal detection data indicates that they also appear more willing to disregard rewards in the face of stress, which is consistent both with the primary reward sensitivity and, at a more speculative level, with the notion that stress is involved in the induction of anhedonia.

2.6 MOTIVATION

A number of studies has found broad cognitive impairments in depression (in both memory and executive-function tests, see Miller (1975); Veiel (1997); Elliott (1998); Austin et al. (1999, 2001); Chamberlain and Sahakian (2004) for reviews; also DSM IV and ICD-10 criteria). Given the breadth of the impairments, some have argued that a generalised motivational deficit may be the most parsimonious explanation (Hasher and Zacks, 1979; Layne, 1980; Schmand et al.,

1994), whereas others (Veiel, 1997) argue more for a global-diffuse impairment of brain function. One finding which has been interpreted as consistent with a motivational deficit is that depressive patients show a deficit on verbal recall, but not on verbal recognition, and that the deficits generally tend to appear for the harder tasks. However, a recent review counts a number of studies which failed to replicate this difference and argues that motivational deficits may only be apparent in severe and / or endogenous cases of depression (Austin et al., 2001).

2.6.1 PSYCHOMOTOR RETARDATION

The most prominent aspect of a general motivational deficit may be psychomotor retardation (PR): slowness and sparseness of thoughts and actions. Clinicians have for a long time related the presence of PR to depression, for example as neurasthenia in the 19th century, in early behavioural formulations of depression (Costello, 1972), or in more recent formulations of endogenous depression (Lemelin and Baruch, 1998) and bipolar depression (Goodwin and Jamison, 1990). For example, motor activity measured over 24h with the aid of a wrist-attached device co-varies with recovery from depression (Wolff et al., 1985). Issues of PR are closely related to the sub-classification of depression into endogenous and reactive types (Nelson and Charney, 1981).

A strong formulation of PR as a distinguishing sign of a subtype of depression is due to Parker and colleagues' work on melancholic depression (Parker and Hadzi-Pavlovic, 1996; Parker and Manicavasagar, 2005; Parker, 2007) who emphasize that PR is a sufficient and necessary clinical feature for the diagnosis of melancholia (a refined notion of endogenous depression). This work merits comment. Parker et al. (1990) analyse a large group of patients diagnosed with varying forms of depression. In a commendable effort to provide a definition of depression in terms of signs (objective features apparent to a clinician observer) rather than symptoms (subjective reports by the patient), they apply factor analysis and principal component analysis to signs of psychomotor retardation found in the patients by a large group (40) of psychiatrists. They find that the first factor (and also the first eigenvector in the PCA analysis) accounts for a large fraction of the variance (40%), and that projection of the data onto this vector divides patients into two groups. The group with the larger projection encompasses most patients diagnosed with melancholic / endogenous depression and characterised by: unresponsiveness to interviewer, dull/inattentive, fixed and immobile facies, slumped posture, immobility, slowed speech and movements, impaired insight and poverty of associations. This forms the basis for their CORE measure, which has better predictive validity for response to ECT and drug therapy than either HamD scores or other measures of endogenicity (Hickie, 1996). Although analysis of the symptoms yielded a less clear classification, they argue that reaction time data on some very simple tasks (Hickie et al., 1990; Rogers et al., 2004) do, but the data presented is not very clear-cut. This may be rectified by different analyses which take the structure of the data into account (for example, the data is bimodal, and this is not respected by either factor analysis or PCA).

2.6.2 BEHAVIOURAL EVIDENCE

There are also tasks more directly reflective of motivation than simple reaction time tasks. For example, in a progressive ratio task the number of responses required to obtain a reward increases after every reward obtained, until the subject terminates responding. This break point is thought to directly reflect the subject's motivation to work for the reinforcer. Hughes et al. (1985) test six subjects with depression before, during and after treatment with antidepressants on a progressive ratio task. The three subjects that respond to treatment (defined as a 50% decrease of the Hamilton Depression (HamD) and BDI scores) work harder for the rewards after recovery (and earn more), while those that do not respond work less (and earn less). There are three complications with this task that prevent us from taking it as strong evidence for a decreased sensitivity to rewards: Firstly, the subject numbers are far too small, secondly, there may again be effects of primary sensitivity; thirdly, there are complex non-stationarities in the task, and patients' ability to make correct inferences about these may interfere with the outcome measure.

A more clearly interpretable test of motivation is response rate on a variable interval (VI) schedule. In a VI schedule, rewards are available at exponentially distributed intervals with a fixed mean, and response rates increase with the size of the available rewards. This task is free of the nonstationarities of the progressive ratio task. Indeed, recent modelling studies (Niv et al., 2005, 2007) show that the response rates in such free operant schedules are related to the average expected reward rate — the motivation — for emitting an action. Szabadi et al. (1981) test two bipolar subjects on a variable interval schedule, In one subject, they find that response rates decrease during hypomanic episodes and increase during manic episodes, returning to a constant level in the episodes of remission. They interpret this as evidence for an alteration in reinforcer sensitivity. Due to the low subject numbers this study counts more as anecdote than evidence at present. Of course, even here, the difference in response rates may still be due to changes in primary reinforcer sensitivity rather than motivation. One approach would be to test the subject in two motivational states. The difference is then informative about the strength of the motivational manipulation. Richards and Ruff (1989) attempt to do precisely this, though not in a free operant scenario. They give subjects a battery of neuropsychological tests and motivated them globally by telling them they will receive $10 upon completion if they do well, and leaving the $10 note in view of the subjects throughout the testing. They find that depressed subjects perform worse, but that there is no interaction with their motivational manipulation (they do check that their manipulation has the desired effects). They conclude that depressed patients may be less motivated (they point out that the defect is not uniform across tasks), but that there is no defect in their ability to be motivated.

2.6.3 EMOTIONAL INDUCTION STUDIES

Conclusions about the involvement of motivational issues in depression have also often invoked emotional induction studies. Reading of emotionally coloured, self-referent statements, leads to increases on questionnaire-indices of depression (Velten, 1968). One caveat is that if motivation is related to expectations of achievable rewards, then it may be directly influenced by statements about general ability, and rather than modelling the effect of depression,

it would be evidence for sensible inference. However, a number of results hint that even here more specific issues may be at large. Raghunathan and Pham (1999) found that subjects chose high-risk/high-reward (uncertain) options after sad mood induction, but low-risk/low-reward (certain) options after anxious mood induction. In the sad mood induction, participants were asked to imagine their mother had died unexpectedly, while in the anxious mood induction, they were asked to imagine their doctor had just called with some important news to divulge, and hints were made towards cancer. While phrased in terms of mood, it is potentially more parsimonious to interpret the induction as instruction to rectify a particular undesirable state. Interestingly, low mood induction has been reported to *de*crease reports of reward sensitivity (the enjoyability of cheese; Willner and Healy 1994), but *in*crease the progressive ratio breakpoint for rewards (cigarette puffs; Willner and Jones 1996). Willner et al. (1998) replicate these effects with chocolate, and find the same in animals: chronic mild stress (see section 2.9.3) in rats increases responding for sucrose solution, despite decreasing the choice preference for sucrose over water. Thus, negative mood induction and chronic mild stress seem to decrease the primary reinforcing value of stimuli in a manner consistent with the evidence reviewed so far, but seem to increase their motivational properties in an operant scenario, which is inconsistent with the (admittedly weak) findings from the progressive ratio studies (Szabadi et al., 1981; Hughes et al., 1985), and with motivational theories of depression (Layne, 1980).

2.6.4 DOPAMINE

It is the tonic, rather than the phasic, aspect of DA that a) is postulated by normative models to relate to motivation (Niv et al., 2005, 2007) and b) has been accessible to experimental assessment in depressed patients. There is now good evidence that changes in tonic nigrostriatal DA function are related to the psychomotor retardation seen in depression. Similar decreases in the mesolimbic system have been postulated to relate to the anhedonia, but this is speculative. We here review evidence relevant to the relationship between tonic DA levels and depression.

Homovanillic acid (HVA) is the main breakdown product of DA and its measurements in depression have yielded relatively clear results. For anatomical reasons, CSF levels of HVA reflect nigrostriatal rather than mesolimbic DA and are tightly correlated with psychomotor retardation. In studies with probenecid[2], HVA levels are consistently decreased, and more so in retarded and bipolar depressed patients (Korf and van Praag, 1971b; Willner, 1983a, 1985b, 2002). It is unclear whether this is cause or consequence (for example, subjects simulating mania show increases in HVA levels; Post et al. 1973), but such a lowered level of DA may be the reason for the compensatory decreases in dopamine transporter densities found by PET (that this is a secondary change is supported by the finding that performance on simple motor tasks is inversely correlated with DAT levels; Meyer et al. 2001) and may account for some of the increases in DA D_2 receptor binding potentials, which may also reflect compensatory changes (Meyer et al., 2001; Shah et al., 1997).

Other data arguing for an involvement of DA in depression are that pro-dopaminergic agents, ranging from DOPA and tyrosine to bromocryptine (a direct agonist) and DA reuptake blockers (nomifensine) do have some anti-depressant activity, mainly in subjects with re-

[2]Probenecid blocks the transport of HVA out of CSF and thereby makes CSF levels more reflectant of DA release

tarded features. They may push bipolar depressed people into a hypomanic state, and have no effect on the agitated or anxious symptoms. Tricyclic antidepressants (TCAs), which affect both DA and other systems on the other hand are helpful in depression both with retarded and with anxious features (Willner, 1985b; Kapur and Mann, 1992; Millan, 2006). Furthermore, psychostimulant drugs have mood elevating effects, and more so in those with higher HamD scores (Tremblay et al. 2002, 2005; though initial reports suggested it may be similar in patient and control populations, see Willner (2002) and references therein). Psychostimulant response predicts efficacy of antidepressant treatment (Little, 1988) and also has direct applications in treatment, mainly in elderly or acute settings (Barbara Sahakian, pers. comm.; Willner 2002).

Whether neuroleptics can induce depression is still controversial. Initial reports that reserpine resulted in depression have turned out to be due to diagnostic errors, but schizophrenic patients on neuroleptics are more likely to show full depressive syndromes than neuroloeptic free patients with equal levels of psychotic symptoms (Harrow et al., 1994). Interestingly, and *prima facie* in full contradiction to these ideas, neuroleptics can also be effective as antidepressants (Robertson and Trimble, 1981, 1982). However, as Willner (1985b) points out, this may be due to a) preferential antagonism of inhibitory autoreceptors (which would still result in a pro-dopaminergic action) and b) may not even work via dopaminergic mechanisms. Nevertheless, neuroleptics are mainly effective in depression with delusional/psychotic features, worsen PR and can also promote depression (Randrup et al., 1975). Finally, administration of the D_2 antagonist sulpiride reinstates symptoms of depression in subjects with a history of depression (Willner et al. 2005; though manipulation of dopamine precursors fails to have any effects; McTavish et al. 2005).

Finally, there is the strong association of depression with Parkinson's disease, which results mainly from dopamine deficiency. Indeed, patients with Parkinson's disease are more likely to show depressive symptoms than controls matched for age, impairment severity and quality of life (Mentis and Delalot, 2005).

2.6.5 MOTIVATION IN DEPRESSION

Thus, certain patients diagnosed with depression do suffer from psychomotor retardation and tend to show evidence of lowered tonic CSF DA levels, while others do not (e.g. the agitated patients). At present, this conclusion is mainly based on observational studies, rather than on direct behavioural assessments of motivation. Behaviourally, the data does not yet differentiate between motivational effects and those of primary reinforcer sensitivity, and the breadth of impairments in depression is still the main argument for a general motivational deficit.

Although it seems important to include this dichotomy in the sub-typing of depression, it is unclear what the specificity of PR by itself is. We are not aware of studies in a community setting testing this. PR can refer to a wide variety of features such as slowness or paucity of movements or thoughts; speech with long latencies and pauses etc., Lemelin and Baruch 1998. It is comparable to what is found in Parkinson's Disease (delays in movements without external cues; Rogers et al. 2000; Sachdev and Aniss 1994; Naismith et al. 2006; see also Parker 2007 who relates melancholia to Parkinson's disease), in schizophrenia (Miller, 1975), and in other situations of decreased attentiveness. All of these diseases have marked comorbidity with de-

pression, and thus it may be that PR, or behavioural measures of it, does describe an underlying commonality, for example one relying on dopaminergic processes. On the other hand, it may also be orthogonal: Lemelin and Baruch (1998) find that the Depressive Retardation Rating Scale (DRRS), which quantifies psychomotor retardation, correlates with performance on several complex attention tasks, but that HamD and CORE do not. While this is not the conclusion of the paper, it may be that measures of retardation such as the DRRS are more apt as a measure of attention than depression.

2.7 THE DEPRESSIVE STATE: RECAPITULATION OF HUMAN EVIDENCE

We have so far reviewed evidence on the state of depression. All the decision making approaches outlined in section 1.2 have been implicated in the depressive state, but this has often been done in a very loose manner and there is as yet very little behavioural evidence.

1. Primary: There is concurring evidence that the state of depression is characterised by a decreased reward efficacy. We would like to stress that this is an overall impression, rather than evidence gained from particular, unequivocal findings. Thanks to the use of pain, the literature on negative reinforcements is more thorough, but also more contradictory. There is strong evidence that the primary effect of punishments (pain) is lessened, but verbal reports show evidence of a magnification. Finally, there is one report indicating that the lack of sensitivity to rewards may be due to the presence of stressors. We will see in the next chapter that such changes to primary reinforcer value are also apparent in animal models of depression, and these will be the focus of chapter 3.

2. Pavlovian: No direct evidence in humans. However, Pavlovian effects can shed light on the relationship between the normal functions of serotonin and its association with depression. This is further explored in chapter 5.

3. Goal-directed: There is evidence that a verbally reported perception of control is important. However, there is at present only indirect behavioural evidence from complex planning tasks. We explore the precise link between control and its putative behavioural sequelae in more detail in chapter 4.

4. Habitual: Overall, there is some evidence that the acquisition of habits may be altered, although at present it is not possible to affirm this as a strong finding independent of that on primary reinforcer sensitivity.

5. Motivation: Findings on motivational aspects are at present not sufficient to draw strong conclusions, but it appears well possible that a subgroup of patients display both motivational deficits and alterations in tonic DA levels.

Overall, this review arguably shows that many of the prominent findings from the three different types of research into depression outlined in section 1.1 (biological, behavioural and cognitive) can be related to each other within affective decision making, and that this therefore

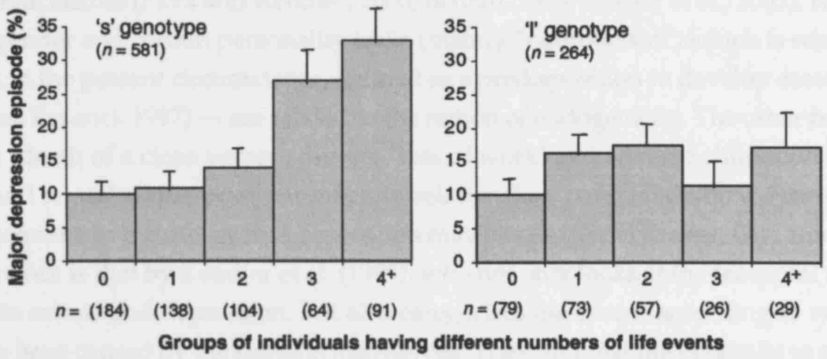


FIGURE 2.4: Interaction of life stress and 5HTTLPR. Subjects from the Caspi et al. (2003) cohort are grouped according to short allele carrier status ('s' genotype contains s/s and s/l individuals, 'l' genotype only *l/l* individuals), and according to the number of life events they experienced. Only in the 's' group do individuals with more life events show higher prevalence of depression. Figure reproduced from Caspi et al. (2003).

provides a useful integrative framework. Furthermore, this view of the evidence suggests a swathe of related, theoretically motivated experiments.

2.8 HUMAN DATA ON THE ÆTIOLOGY OF DEPRESSION

Evidence pertaining to the function of various decision making subsystems during the depression was reviewed in the previous chapter. We now proceed to review evidence on the induction of the depressed state. The most prominent known inducing factor is stress (Kessler, 1997). In humans, evidence is mainly based on epidemiological findings, though there are some psychological manipulations which speak to the same issue. The majority of our understanding derives from animal models of depression, which will be the concern of the next section.

2.8.1 STRESS AND GENETICS

Nevertheless, two general accounts in aetiological investigations into depression stand out, and have existed throughout the latter half of the past century. First are the endogenous depressions (also called endogenomorphic or melancholic, and partially reviewed above), which are assumed to arise from within a subject, probably via a genetic predisposition. In contrast to these are the reactive (neurotic) depressions, which are assumed to be kicked off by *stressful external* events (Kessler, 1997). To some extent, this was based on a division of patients' reports: some clearly reported very stressful events, other did not. Careful analyses and questioning revealed early on that this was due to a reporting bias amongst those who had apparently not experienced stressors (Akiskal and McKinney, 1973; Kasper et al., 2003), but the distinction survived. In support of these two aspects, studies on large cohorts have consistently identified

four main risk factors (Fava and Kendler, 2000; Bentall, 2003; Kasper et al., 2003), two of which — female gender and certain personality traits (mainly "neuroticism", which is somewhat tautologically, in the present circumstance, defined as a predisposition to develop emotional upset under stress; Eysenck 1997) — are related to the notion of endogenicity. The other two, stressful live events (death of a close person, divorce, loss of work) and adverse childhood experiences (physical and sexual abuse, poor parent-child relationship, parental divorce; Heim et al. 2000) are more germane to the theory that depression may have external causes. One study of particular importance is that by Kendler et al. (1999), who not only looks at the temporal relationship between life events and depression, but also categorises life events according to whether they might have been caused by the subjects themselves. They find that life events have a substantial causal role in the onset of depression, although it should be pointed out that the relationship between stress and depression is not purely unidirectional, as persons suffering from depression self-select into high-risk environments (Anisman and Matheson, 2005; Kendler et al., 1999); that past episodes of depression may increase the effect of milder stressors relative to severe ones (Kendler et al., 2000); and that depressed people may have worse social skills (see Williams (1992), chapter 2).

Recently, the two concepts have been combined. Kendler et al. (1995) pioneered an analysis which looked at the interaction of environmental and genetic risk factors. Mono- and dizygotic twins were categorised according to whether they a co-twin suffered from a mood disorder (high-risk) or not. They found that there was a genetic predisposition for developing depression after life events (four of these had an odds ratio > 10 for depression in the months following them: death of close relative, assault, serious marital problems and breakup/divorce). Low-risk genetic groups had probabilities of MDD onset of 0.5% without and 6.2% after serious life event, while for the high-risk groups it was 1.1% and 14.6% respectively. So it seems that genetic predisposition modifies the sensitivity to life events. Based on results in primates and rodents (Murphy et al. 2001a; Bennett et al. 2002; Barr et al. 2003, see also Leonardo and Hen 2006), Caspi et al. (2003) succeeded in identifying the 5HT transporter as a gene that might underlie this gene×environment interaction (Caspi and Moffitt, 2006) in humans. Figure 2.4 shows that only in individuals with the less effective, short allele of the 5HT transporter gene an increase in the number of life events translates into an increase in the prevalence of depression. This is an extremely exciting finding. It has been hotly debated and replications have mostly been successful (Eley et al., 2004; Kaufman et al., 2004; Gillespie et al., 2005; Grabe et al., 2005; Kendler et al., 2005; Wilhelm et al., 2006). Importantly, Kendler et al. (2005) find that the polymorphism mainly has an effect on the development of MDD after mild stressors, and that its presence specifically affects the development of depression but not that of anxiety. However, we should keep in mind that stress leads to depression even in those with the protective *l/l* alleles (Kendler et al., 2005). In another study on women, Kendler et al. (2000) find that the odds ratio for the development of depression is 10 after the first life event — irrespective of the genetic status.

When considering the induction of depression, we will neglect the effect of genes in chapter 3 and concentrate on the effect of stressors by themselves. We will give some explanations for why depressed people may fail to avoid stressful situations, and how this may contribute to the maintenance of depression in chapter 5 and 4. Chapter 4 will contain some results that may

relate to genetic predispositions.

2.8.2 CONTROLLABILITY

We saw in section 2.4.1 that there is some evidence that the LH paradigm can a) be recreated in humans and b) that such induction of helplessness might be associated with lowering of mood. However, we also reviewed a number of major caveats. More decisive data comes from epidemiological settings, in which negative attributional styles increase the likelihood of developing depression, and normalise upon recovery (Haffel et al., 2005). For example, Alloy et al. (1999) followed two cohorts of undergraduates for 5 years. The cohorts were selected from over 5000 undergraduates in terms of scores that assessed how internal, global and stable their attributions for negative events were. The Cognitive Style Questionnaire, an expanded version of the Attributional Style Questionnaire (Peterson et al., 1982), was used to identify attributional styles and the Dysfunctional Attitudes Scale (Weissman and Beck, 1978) was used to identify peoples' schemas. A subset of those with high and low scores were then followed up. Amongst those in the high-risk group, 17% went on to develop depression, whereas this occurred only for 1% in the low-risk group. As Bentall (2003) points out, this still implies that most cases of depression are missed because only 1.3% of subjects were in either of the two groups. It is also not clear whether the low-risk group does better than the rest of the population, i.e. whether expectations of no hopelessness may contribute towards resilience (Maier et al., 2006; Barnett et al., 2006). Lyon et al. (1999) added further credibility to these conclusions by comparing an explicit test (ASQ) and an implicit memory test, in which subjects were given positive and negative, self-referent statements which they had to remember and answer questions about. In both tests, depressed subjects made more global, internal and stable attributions for negative events than controls. Thus, people who consciously *report* little personal control to acquire rewards, and little personal control to avoid punishments are more likely to develop depression.

2.9 ANIMAL MODELS OF DEPRESSION

> Model airplanes do not need to make transatlantic flights; they only need to embody the essence of flying in an airplane (Martin EP Seligman, in Peterson et al. 1993).

2.9.1 VALIDITY

Many psychiatric phenomena cannot, and will never, be captured by animal models because of the absence of culture, language and differences in consciousness. But its weaknesses may be its strengths, because animal work is not confounded by these hard-to-control issues, and can benefit from a century's worth of behavioural animal work. Importantly, animal work can employ invasive and stressful experimental designs and thereby probe the neurobiology of diseases much more thoroughly.

Animal models of depression come in all colours and shapes, from genetic, to behavioural,

lesion, pharmacological and developmental (Barr et al., 2003; Willner and Mitchell, 2003; Cryan and Mombereau, 2004; Anisman and Matheson, 2005; Cryan and Holmes, 2005; Frazer and Morilak, 2005). We will here only be concerned with the two animal models that are phrased in the most behavioural terms and relate directly to affective decision making: learned helplessness and chronic mild stress (CMS). The precise degree to which animal phenomena in LH and CMS model depression has been debated extensively (McKinney and Bunney, 1969; Willner, 1985b, 1986, 1997; Willner and Mitchell, 2002; Frazer and Morilak, 2005; Dulawa and Hen, 2005; Maier and Watkins, 2005) and we refer the reader to these discussions for details. Briefly, recent discussions concentrate on:

- Face validity: does the animal model mirror the array of symptoms and signs in the human disease?
- Construct validity: is there a theoretical rationale to the animal model that yields insights into the human condition?
- Predictive validity: are the behavioural signs in the animals sensitive to and relieved by drugs that are efficient in a clinical setting?
- Discriminant validity: how specific a model is it? Does it differentiate between related disorders (e.g. depression and anxiety)?

Learned helplessness generally shows all these aspects of validity, but to a lesser degree than CMS. Perhaps the most glaring lacuna of LH helplessness as a model of depression is its lack of discriminant validity with respect to anxiety, though a similar issue may be present (to a lesser extent) for CMS (Strekalova et al., 2004).

In both cases, a state reminiscent of depression (in terms of the various validity criteria) is induced by stressing the *normal (!)* animal, and terminated by drugs and treatments (even ECT and exercise) that are efficient in human depression. This is worth emphasizing: the putative depressed state is induced in *normal* animals. It is *not* confined to "depressed" animals. To that extent, it is the "right", normative response to uncontrollable stressors. Indeed, LH in animals arose from investigations into two-factor theories of learning, and it has been investigated extensively for its own sake as a learning phenomenon of interest, not only also as a model of depression (to a lesser extent this is also true of chronic mild stress). Therefore, despite their limitations in reproducing the exact *state* of depression, models involving induction of depression by stress are of particular interest to us because a) they form a link to normative decision making and b) make clear, testable claims about the computational mechanisms involved, and the kinds of state that may be reached by the interaction of the environment with a normal, functional affective system. We iterate again that the huge prevalence of depression to us signifies that, at least in a large fraction of cases, there must be signs of normal mechanisms at work. As mentioned before, one hypothesis we would like to approach in this thesis, is whether some forms of depression can result from maladaptive interactions between essentially normal decision making systems that work in parallel. Before we proceed with this, the interaction of the systems has to be understood. It is in the clarification of this point that we view the animal work as crucial and highly informative.

2.9.2 LEARNED HELPLESSNESS

In their seminal paper, Overmier and Seligman (1967) found that dogs which had experience of inescapable electric shocks failed to learn an avoidance response, and did not even bother escaping once the shock had come on. Seligman and Maier (1967) introduced the triadic design which controlled for shock exposure and isolating the effect due to action-outcome contingency (see figure 2.2 and section 2.4.1). It is the perceived action-outcome contingencies that matter: helplessness in the escape task is not observed if the yoked rats are also supplied free-running wheels, as they will very actively turn the wheels during shocks and thereby experience the same action-outcome contingencies as the master rats (Steven Maier, pers. comm.), at least over the timescale of the experiment. Maybe for this reason the transfer of the paradigm from dogs to rats presented a challenge: only when the shuttle box escape response consisted of two crossings (back and forth), was there a consistent helplessness effect after IS, and arguably the escape deficit only appears when the action is not "natural" to the animal (Seligman, 1975).

Claims that the deficits are due to a perception of lack of control were supported by its generality across environments and reinforcers. The very first demonstrations of the effect already included the generalisation across environments (from Pavlov's hammock to the shuttle box), though right from those early papers, there are descriptions of the nontrivial and not obviously normative dynamics of this generalisation: animals initially escape correctly, but then start to fail and do not acquire the escape response (Overmier and Seligman, 1967; Maier and Watkins, 2005). There is a strong consensus that LH generalises across stressors. For example, rats also showed deficit for escape from a flooded alley or a water maze after IS (Maier and Seligman, 1976; Lee and Maier, 1988)[3] and escape deficits are apparent when the escape action consisted of a *de*crease in activity (rats had to stay still to escape the shock; Willner 1983b; Seligman 1975). Furthermore the effects seem to be bidirectional (Mineka and Hendersen, 1985): IS rats show some facilitation for learning a non-contingency (Testa et al., 1974), while exposure to ES delays acquisition (Volpicelli et al., 1983). While most studies find that ES rats and controls do not differ, some do find that prior experience with control (ES) does facilitate acquisition of an escape response (Goodkin, 1976). This bi-directionality is consistent with the notion that animals infer some quantitative measure of control, or action-outcome contingency, based on the entirety of their experiences in the relevant environment and use this to plan further actions.

Strong evidence that the escape deficit is really due to a lack in the contingency perception is due to Jackson et al. (1978). Here rats are first trained on an appetitive lever-press task. After reaching criterion, rats are exposed to ES or IS, and returned to the appetitive lever-press chamber. Rats are then divided into two groups. One group is given a simple suppression task: during a CS, they are given two shocks, independent of their lever presses. The other group is only given shocks during the CS if they respond on the lever. The authors report differences between IS and ES rats only on the contingent task, and not in the non-contingent shock scenario. We will discuss this evidence in great detail in chapter 3.

Although there has been comparatively little emphasis on this aspect, there is rather consistent evidence that LH generalises not only to conflict scenarios, but across reinforcer valence

[3]Part of the effect may however be specific to the inescapable stressor, or to the context in which the stressor was experienced (Minor and LoLordo, 1984), which is important as it may be due to learned irrelevance (Mineka and Hendersen, 1985).

(Job, 2002). Goodkin (1976) and Welker (1976) report that operant control of food (or shock) enhances subsequent escape avoidance learning in rats and pigeons respectively. Others report that IS interferes with the learning of an appetitive task (Rosellini, 1978; Goodkin, 1976; Overmier et al., 1980), but interestingly not with the maintenance of an appetitive task learned prior to the exposure to IS (Ghiglieri et al., 1997; Mangiavacchi et al., 2001). Similar to the Volpicelli et al. (1983) study above, Rosellini et al. (1982) train rats on an appetitive lever-press task and then gives IS or ES. Rats are then returned to the lever press cage, but food delivery made non-contingent. Rats persist less if previously exposed to IS. There have also been attempts to characterise what was termed a "cognitive" deficit in rats (Jackson et al., 1980), but the consistent initial reports were due to confounds of an attentional nature (Minor et al., 1984). This will be discussed in more detail in chapter 3 and form part of the motivation for chapter 4.

Amat et al. (2005) make an anatomical point which is extremely intriguing, as it begins to address questions about the goal-directed and habitual nature of the effects. The authors note that the dorsal raphé nucleus (DRN) is under inhibitory control by the ventromedial prefrontal cortex (PFC). Inhibition of the DRN prevents the effects of IS (Maier and Watkins, 2005). Thus, inhibition of the vmPFC prevents the protective effect of control: master rats with inhibited vmPFC show the same behavioural deficits as yoked rats. While this is a fascinating result, in the light of the findings by Killcross and Coutureau (2003), it would be extremely valuable to know whether the suppression of the DRN after IS is implemented by the goal-directed, or the habitual system.

There are major caveats. We have seen that the central concept in LH, control, is about estimates of a reinforcer's probability, not its size. But rewards themselves interact: both shocks and rewards affect subsequent reinforcers. Pain-induced analgesia is as common as pain-induced hyperalgesia (Kelly, 1986). Feeding influences the nociceptive response (Fontella et al., 2004), but this is modulated by stress and even varies for different tastes (de Vasconcellos et al., 2006). Pain-induced analgesia can be opioid and non-opioid (Terman et al., 1984; Maier, 1989), and depending on this can have varying effects on rewards.

However, the effects of IS and ES on the sensitivity to reinforcers themselves appear different: IS, unlike ES, results in a shift of the intracranial self-stimulation (ICSS) dose-response curve such that IS animals are less sensitive to electrically induced DA release (Zacharko et al., 1983; McCutcheon et al., 1991), and IS animals show deficits in other behaviours that are mediated by DA, such as avoidance learning (Friedhoff et al., 1995). Jackson et al. (1979) show that analgesia only occurs after IS, while there is no change in (shock-induced) pain thresholds after ES and Grau et al. (1981) show that ES before IS prevents analgesia (at a similar efficacy to morphine).

Finally, there is also fear, which is a subtle issue (Maier and Watkins, 1998). Fear might be conceptualised as the expectation of punishments: future punishments are given a large probability. This might fit in nicely with helplessness, for example because punishments cannot be escaped. Fear affects pain (Vowles et al., 2006) and conditional analgesia can happen in response to contextual cues Amit and Galina (1986); Lysle and Fowler (1988), and non-noxious events (Lee and Rodgers, 1990). While IS is inevitably associated with fear and anxiety (Maier and Watkins, 2005), just as depression is, it seems that the escape deficit itself does not depend on fear: anxiolytics during the escape task have no effect (although they do abolish LH when

given during IS Drugan et al. 1984). Maier et al. (1993) do a LH experiment but measure fear in the shuttle box. They find a double dissociation: amygdala lesions impair the fear conditioning but have no effects on LH, while DRN lesions abolish LH but leave the fear conditioning intact.

Thus, animals note the different between controllable and uncontrollable situations, but it is not clear that animals fail to learn tasks because of a belief that outcomes are independent of action. The evidence on escape deficits after IS so far indicates that in situations in which there is no control, animals reduce their sensitivity to reinforcers. This may or may not rely on inference of control. We will see that a stronger case for the importance of control comes from the generalisation across reinforcer valence.

2.9.3 CHRONIC MILD STRESS

A number of aspects of LH cast doubts on its validity as an animal model of depression: First, escape deficit is not a prominent feature of depression; second, the effects of IS only last for approx. 48 hours (though see Overmier et al. 1980; Maier 2001), and the antidepressant medications exert their effects within this time (unlike antidepressants in humans which take up to several weeks to show effects); third the shock procedure is more akin to a traumatic event than to a life event, and so may resemble PTSD more than depression (Maier and Watkins, 2005).

Chronic mild stress (CMS) is a more recent animal model developed to address these three points (Willner et al., 1987). It has extremely good predictive, face and construct validity (Willner and Mitchell, 2003; Willner, 2005). Rats are exposed to a whole series of mild, variable stressors over a period of weeks to months. In its mildest versions, the stressors include cage tilt, wet bedding, continuous lighting and food and water deprivation, but more severe forms including (few) tail shocks, and tail suspension are now frequently used mostly to ensure more reliable effects. These mild stressors are more similar to life events: for example, unemployment or divorces bring a myriad of daily small stresses.

The standard outcome measure of CMS is preference of dilute (1%) sucrose (or the non-caloric saccharin) over water. While unstressed rats mostly choose the sweet solution, stressed rats show no preference for a period of several weeks following termination of the mild stressors and preference is restored by chronic as opposed to acute antidepressant treatment. Not preferring palatable sucrose over water is argued to resemble human anhedonia closely (Willner et al., 1987; Willner, 1997, 2005). Chronically stressed rats also show decreased place preference (Papp et al., 1992), increased intra-cranial self-stimulation (ICSS) thresholds (Moreau et al., 1992) and decreases in spontaneous motor activity (Grippo et al., 2003). There is good evidence that the decreased sucrose preference goes hand-in-hand with impaired appetitive learning on the Y-maze (Ghiglieri et al., 1997; Gambarana et al., 1999), although there have been reports that both appetitive and consummatory responses for sucrose are decreased without a corresponding decrease in progressive ratio schedule break point (Phillips and Barr, 1997; Barr and Phillips, 1998; Willner et al., 1998), and even with an increase (Willner et al., 1998).

Just as IS, CMS is reported to have symmetric effects on positive and negative reinforcers: it increases anxiety on most standard measures (Strekalova et al., 2004), and impairs aversive learning (shuttle-box escape; Murua et al. 1991; Gambarana et al. 2001). To our knowledge, pain

sensitivity has not been tested explicitly in CMS, but reports that even very mildly stressful experiences can induce analgesia (Mogil et al., 1996; Lee and Rodgers, 1990) would make this very likely. Intriguingly, there are also two reports of asymmetric effects: Papp et al. (1992) report that CMS affects appetitive but not aversive place conditioning, and Harding et al. (2004) find that CMS decreases responding for a signalled food reward without affecting responses for a signalled shock. In the latter study, rats did not show decreased sucrose preference, indicating that CMS might have affected the habitual system (the learned response) directly.

In CMS, there is no "master" rat with control, and due to the difficulty and lengthy nature of the experiments, there has been very little parametric exploration of the stressors (Willner, 2005). However, it appears that the mild stressors have to be both unpredictable and varied (unlike more severe Maier 2001; Ghiglieri et al. 1997), and otherwise only produce short-term effects which habituate, and may even induce patterns opposite to depression (Muscat and Willner, 1992; Cabib and Puglisi-Allegra, 1996).

Unlike in LH, the emphasis in CMS is on alterations of the reward process itself. While there are two interesting reports suggesting this is selective and does not apply to punishments, they weight of evidence favours a symmetric effect that also extends to impaired punishment processes. CMS data has been interpreted as leading to a change in primary reinforcer (reward) sensitivity, rather than a change in the judgement of outcome probabilities. Nevertheless, we saw that a similar argument could also have been made for the LH data: there too, reward and punishment processes were blunted. We conclude that controllability may play a similar role in CMS as it did in LH, although there is as yet very little evidence to bear on the issue.

2.9.4 Neuromodulators

2.9.4.1 Dopamine

Dopamine function is generally decreased in animal models of depression, either as a direct consequence of manipulations of the dopamine system (Barr et al., 2002), or, more pertinent to induction, as a consequence of stress. We have already seen that the reduction of dopamine function has directly been linked to anhedonic features of the models and go hand in hand with decreased sensitivity to rewards and impaired acquisition of appetitive tasks.

Dopamine responds mainly to events of rewarding nature, but acutely, aversive events also lead to local and distal DA release as measured both by microdialysis and voltammetry in response to a wide variety of stressors, such as foot shock, tail pinch or restraint stress (Deutch et al., 1990; Imperato et al., 1991; Cabib and Puglisi-Allegra, 1996; Horvitz, 2000; Ungless, 2004). The increase appears not to be mediated by increased DA neurone firing; extracellular recordings in the VTA indicate that activity is generally suppressed by stressors (Ungless et al., 2004; Ungless, 2004; Mirenowicz and Schultz, 1996); but because the suppression is not total, there is ample scope for modulation of DA terminal boutons, for example by 5HT. When the stressor becomes chronic rather than acute, a variety of adaptations are seen. Cabib and Puglisi-Allegra (1996) suggest that chronic exposure to stressors that are not under behavioural control of the animal lead to a blunting of the dopaminergic response to stressors, whereas it is maintained if there is behavioural control. In fact, inescapable and escapable shocks produce the same DA

responses in the NAcc (Zacharko and Anisman, 1991; Cabib and Puglisi-Allegra, 1994; Bland et al., 2003b), although some metabolite measurements suggest that DA release might be decreased by IS and increased by ES (Cabib and Puglisi-Allegra, 1994). In the PFC, IS results in a larger DA release than ES (Bland et al., 2003a). The high levels of prefrontal DA may induce adaptive processes that lead to subsequent decreases in DA responsiveness. When looking at the time course of DA levels over one prolonged episode of (mild) stress, there is an initial increase followed by a decrease (Imperato et al., 1991; Puglisi-Allegra et al., 1991). If the stress is repeated over a few days, the time course shifts downwards: there is no more positive response at the beginning of the stress episode, and the subsequent decrease is more pronounced (Imperato et al., 1993). In fact, chronic stress pushes the tonic levels of dopamine down at all times, not just in periods of stress (Gambarana et al., 1999; Masi et al., 2001). As a caveat we note that DA responses in both NAcc and PFC to new stressors may be increased after CMS (Di Chiara et al., 1999).

The trend of a reduction in responsiveness after stress extends from punishments to rewards. Severe stress as used in the LH paradigm increases ICSS thresholds in the limbic but not the motor part of the dopaminergic system (VTA, NAcc, medial PFC; Zacharko et al. 1983; Zacharko and Anisman 1991). CMS-induced increases in the ICSS threshold have also been documented (Moreau et al., 1992). Voltammetric measurements have confusingly found that CMS increases the amount of dopamine released when the VTA is stimulated electrically (Stamford et al., 1991), but microdialysis measurements after (more severe) stress have found decreased release by both amphetamine (Gambarana et al., 1999) and by appetitive events (Di Chiara and Tanda, 1997; Di Chiara et al., 1999; Gambarana et al., 2003). This diminished DA release may be compounded by D_1 and D_2 receptor adaptations (Willner et al., 1991; Papp et al., 1994; Scheggi et al., 2002).

It is clear that some of the behavioural sequelae of stress on both rewards and punishments are mediated by dopaminergic changes. Thus, CMS induces a subsensitivity to DA agonists (Papp et al., 1993a) and DA sensitisation by amphetamine reverses the deficit in sucrose preference seen after CMS (Papp et al., 1993b). A similar effect is obtained after behavioural stimulation of the appetitive system: rats that are first trained on an appetitive discrimination task and then exposed to chronic stress do not show impairments in the appetitive tasks. Their comrades, which are first exposed to the stress, fail to acquire the appetitive discrimination response (Ghiglieri et al., 1997). The latter set of rats develop a subsensitivity to amphetamine, while the former does not (Masi et al., 2001). Gambarana et al. (2003) push these findings one step further, arguing on pharmacological grounds that stress does affect the phasic DA response. They show that chronic lithium leads to a tonic DA decrease as chronic stress does, but that it does not impair acquisition of the appetitive discrimination task and it does not affect the phasic (though microdialytic) DA response to appetitive events. Finally, even the aversive shuttle box escape deficit in LH is dependent upon changes to dopamine: chronic imipramine leads to a dopaminergic hypersensitivity mediated by D_2 receptor up-regulation which LH after IS and is itself inhibited by D_2 antagonists (Gambarana et al., 1995b; Besson et al., 1999; D'Aquila et al., 2000; Naranjo et al., 2001), D_1 antagonists can reverse the protective effects of imipramine and D_1 or D_2 agonists alone can acutely overcome the effect of prior IS on escape performance (Gambarana et al., 1995b). D_1 receptor desensitisation can itself also produce an

escape deficit (Gambarana et al., 1995a).

2.9.4.2 SEROTONIN

In near perfect opposition to the indoleamine hypothesis (section 2.3.1; Lapin and Oxenkrug 1969), animal models of depression generally find that the behaviours thought to reflect depression are accompanied by *enhanced* 5HT function. However, it is unclear whether this really reflects processes central to depression (such as anhedonia), or rather anxious processes, which accompany all behavioural models of depression. It is also unclear whether regional specificities in 5HT receptor distributions contribute to the picture, and how these changes relate to the induction as opposed to the state of depression. We will here review animal data on four very specific issues that are in no way representative of the wide array of functions ascribed to serotonin:

1. Functional role of 5HT in animal models of depression
2. Effect of antidepressants (in models of depression or separately)
3. Effect of 5HTT on 5HT levels, anxiety and depression in animals
4. Interaction between 5HTT and stress

In terms of its functional role, increases in DRN 5HT are argued to be both sufficient and necessary for the behavioural deficits after IS (Maier and Watkins, 2005). The two fundamental observations in support of this interpretation are:

- DRN 5HT neurones are far more activated by uncontrollable stress than by controllable stress (*c-fos* measurements; Grahn et al. 1999b; Takase et al. 2004, 2005). Although this activation does not last as long as the effects of of IS, new shocks are more efficient at activating the DRN for several days, paralleling the length of the behavioural effects of IS (Maier and Watkins, 1998). Levels of 5HT in projection areas of DRN, such as the basolateral amygdala, PAG and the vPFC are thereby increased (Amat et al., 1998a,b; Bland et al., 2003a).
- Inhibition of this DRN response by lesion or pharmacology blocks behavioural sequelae of IS (Maier et al., 1993, 1995b) and, conversely, excitation of DRN 5HT neurones reproduces the effects (Maier et al., 1995a). Grahn et al. (1999a)'s experiments demonstrate this quite elegantly. DRN neurones express an inhibitory μ-opioid receptor. Injection of morphine prior to IS potentiates the effect, while injection of naltrexone antagonises it.

Thus, 5HT in LH is argued to principally mediate (but the raphé is not itself argued to compute) the effects of *control*, rather than other effects such as analgesia. However, while the LH effects certainly do generalise to appetitive learning scenarios, we are not aware of data showing that the sequelae of these serotonergic manipulations do, and the data does not show that 5HT is involved in the mediation of the anhedonic aspects of LH. Indeed, DRN 5HT is involved in the mediation of the analgesia after IS (Sutton et al., 1997) and particularly the serotonergic facets

of LH may be more closely related to anxiety than depression (Maier and Watkins, 1998, 2005; Willner and Mitchell, 2003). In fact, a number of prominent theories of anxiety are centred around serotonin's anxiogenic properties (Deakin and Graeff, 1991; Graeff et al., 1998; Gray, 1991), particularly that reductions of its raphé levels release punishment-suppressed responding (Graeff, 2002) and produce impulsive behaviour (Soubrié, 1986; Fletcher, 1995; Mobini et al., 2000a). The relationship of CMS, which focusses more on anhedonia than LH, to serotonin is less thoroughly explored. There are reports of $5HT_{1A}$ adaptations, but unlike in LH, this is accompanied by a generalised decrease in 5HT and 5-HIAA levels in brainstem, striatum and cortex (Lanfumey et al., 1999; Grippo et al., 2005), despite the reported generality of anxious behaviour after CMS (D'Aquila et al., 1994; Strekalova et al., 2004). Thus at present it is unclear whether the serotonergic effects really have to do with the anhedonic aspects of the animal models, or whether they are more prominently related to the pervasive anxious effects.

SSRIs act on 5HT by inhibiting its reuptake, which increases extracellular concentrations of 5HT. Despite the rapid pharmacological effect, the therapeutic effect is delayed. In rodents, the therapeutic effect is paralleled by an adaptation of the inhibitory $5HT_{1A}$ autoreceptors on raphé neurones. Removal of this autoinhibition in combination with reuptake inhibition is argued to increase 5HT levels in distal projection regions slowly over a few weeks (Willner, 1985a; McAllister-Williams and Tyrer, 2003; Millan, 2003; Hariri and Holmes, 2006). Surprisingly, LH does respond to subchronic (3-5 days) serotonergic antidepressant treatment, via either SSRIs or selective agonists at $5HT_{1A}$ receptors (Willner and Mitchell, 2003). Thus, 5HT boosts are central to the LH after shocks, but further 5HT increases alleviate the effect. Indeed, the mechanism by which LH has been hypothesised to increase DRN 5HT neuronal activity is precisely by the same adaptation of $5HT_{1A}$ receptors (Laaris et al., 1999; Maier and Watkins, 2005). Given its reported 5HT decreases, it is maybe less surprising that CMS responds to SSRIs and a number of other atypical serotonergic agents (Muscat et al., 1992; Willner and Mitchell, 2003). Throughout, however, we are not aware of any evidence that the direct effects on 5HT of SSRIs (or any other serotonergic medications) is responsible for the reversal of the anhedonic effects. Rather, as we saw above, it is likely that these effects are due to indirect actions at DA receptors.

What about the allelic polymorphism in the 5HT transporter gene? We saw that its interaction with stress is associated with depression (section 2.8.1), but in absence of the stress factor, the short allele is associated with anxiety (Lesch et al., 1996). The same is the case in animal studies. Although there are obvious developmental caveats to knock-outs (e.g. Ansorge et al. 2004, and even Zhao et al. 2006 who report a living total 5HT knock-out), two independent 5HTT functional knock-out mice show strong anxiety-related traits (Hariri and Holmes, 2006) and also some depressive features, such as increased floating in the forced swim test (Lira et al., 2003). The precise picture may depend on the receptor, as a $5HT_{2A}$ knock-out showed mainly anxious behaviours (Weisstaub et al., 2006), but $5HT_{1B}$ and $5HT_{2C}$ knock-outs enhanced behavioural responses to cocaine (Rocha et al., 1998, 2002). Furthermore, it may be that, as in humans, the interaction with stress leads to further depressive features. An indicator that this might be the case comes from rhesus macaques, who have a homologue of the human 5HTTLPR polymorphism which interacts with rearing environment and influences CSF 5-HIAA levels, aggressiveness and willingness to engage in social play with peers (Barr et al., 2003,

2004).

2.9.4.3 DOPAMINE - SEROTONIN INTERACTIONS

Thus, serotonin and dopamine appear to have dissociable roles in depression and in animal models of it. While dopamine relates more strongly to anhedonia and psychomotor retardation, serotonin by itself confers anxious features, but it also appears important in the stress diathesis.

But they interact significantly and are not independent entities at all. Dopamine and serotonin have long been seen as opponents (Daw et al. 2002 and references therein). This notion rests mainly on behavioural evidence and was given added weight by the hypothesis that non-classical antipsychotics alleviated some of the extrapyramidal side-effects and helped in the treatment of the negative (anhedonia-related) symptoms by action on $5HT_2$ receptors (Kapur and Remington 1996, 2001 though see also Kapur and Seeman 2001). Recently, 5HT agonists or SSRIs have been used to combat obesity and other addictive, putatively DA-mediated, behaviours (Higgins and Fletcher, 2003). Thus, in rodents, 5HT antagonises the general arousing effects of DA (Carter and Pycock, 1978), the self-administration of amphetamine (Higgins and Fletcher, 2003) and ICSS (Redgrave, 1978), the effects of dopamine on appetitive learning (Fletcher, 1996) and the potentiation of appetitive learning by amphetamine (Fletcher et al., 1999). Given the huge variety of receptor types, it is not surprising that there are variations on this theme, but the $5HT_{2C}$ receptor has yielded the most consistent data on may account for much of the consistent opponency of 5HT to DA (Higgins and Fletcher, 2003).

However, in opposition to these findings is that serotonin increases dopamine levels directly in the NAcc and other structures (Parsons and Jr, 1993; Galloway et al., 1993; Fletcher et al., 2002; Benaliouad et al., 2006), maybe due to competition at the DA transporter (Sulzer and Edwards, 2005); and that the rise in 5HT due to SSRIs has overall pro-dopaminergic effects, in terms of behavioural and physiology (Serra et al., 1979; Besson et al., 1999; D'Aquila et al., 2000; Sasaki-Adams and Kelley, 2001) Indeed, there is one report that DA antagonists reverse the antidepressant effect of SSRIs (Willner et al., 2005). Furthermore, SSRIs and other antidepressant are also efficient in the treatment of anxiety, although the more prominent animal models of serotonin and anxiety more readily predict that a decrease in serotonin should be anxiolytic (Graeff, 2002). We will briefly return to the interplay between the two neuromodulators in chapter 5, though this evidence is mainly mentioned as a caveat against too simplistic an attribution of independent roles to the two neuromodulators.

2.10 INDUCTION: RECAPITULATION

Thus, stress is the main known aetiological factor in the induction of depression in previously healthy humans. Stress can also induce depression-like states in normal animals, and these are relatively well characterised in terms of affective decision making and their neurobiological underpinnings.

In humans, the data is not sufficient to argue for a preferential involvement of any particular decision making system in the onset. In animals, stress appears to change the subsequent

primary sensitivity to both rewards and certain punishments. There has been little direct investigation of the impact of stress on goal-directed behaviours (such as motivational shifts; Dickinson and Balleine 2002), but the importance of the prefrontal cortex (Amat et al., 2005) and the fact that learning impairments are present after uncontrollable presentation of reinforcers of both valences argue for such a mechanism. We will show that, computationally, control is related in important ways to motivation, although the neurobiological correlates of this are as yet unexplored. In animals, there is no behavioural evidence for a contribution by an impairment of the habitual learning mechanism, although some of the evidence on relatively phasic reponses of DA to rewards does hint that this might be the case. Finally, the involvement of serotonin is very complex. It mediates the effects of uncontrollable stress in the acute scenario, but the relevance of serotonergic activations to the effects of chronic stress and to the generalisation across reinforcer valence (the best indicator for an involvement of goal-directed control) is not clear. We will see that some of the apparent contradictions around serotonin can be understood through Pavlovian action choice.

In the chapters to follow, which contain the body of the thesis, we present an initial attempt to clarify such issues for some aspects of the animal data. This work is undertaken both because the animal behaviour is of interest in itself, but more importantly in the hope that it will facilitate a similar dissection in human behavioural paradigms.

III

BLUNTING

ABSTRACT

From its origins as a habitual learning experiment, learned helplessness (LH) has come to be one of the conceptual cornerstones of research on depression. In LH, animals that are exposed to inescapable, uncontrollable shocks develop analgesia, unlike those exposed to shocks that are under behavioural control. Blunted pain sensitivity is also a feature of human depression. Here, we use computational tools to ask two questions: firstly, what are the learning consequences of having access to analgesia, i.e. being allowed to change the primary reinforcer strength? Secondly, what aspects of the LH data can the combination of analgesia and cached, habitual learning — which does not rely on an explicit measure of control — account for? We find that analgesia has profound consequences on the ability to shift policies, and that a number of key aspects of the data are well-accounted for by models that make no recourse to any explicit notion of control. We also introduce important notions of generalisation.

3.1 INTRODUCTION

Learned helplessness (LH) has without doubt been the most influential model in research on depression (Maier and Seligman, 1976; Depue and Monroe, 1978; Abramson et al., 1978; Willner, 1985b, 1986; Wong and Licinio, 2001; Willner and Mitchell, 2003; Cryan and Holmes, 2005). Despite its numerous limitations (Costello, 1978; Willner, 1986; Willner and Mitchell, 2003; Maier and Watkins, 2005; Frazer and Morilak, 2005), the model is still of great importance, in both its human and animal incarnations. The human, cognitive, aspects of LH have given

important theoretical credence to the development of non-pharmacological therapies (Willner, 1985b; Williams, 1992; Bentall, 2003; Woolfe et al., 2003). However, animal work has been invaluable in allowing both a detailed analysis of the underlying neurobiology and pharmacology, but also an in-depth dissection of its psychological components. Because human data has usually lacked the clear behavioural outcome measures and has been marred by confounds of verbal reports, the psychological analysis of the concept of control has been more convincing in animals than humans. Despite their greater distance from the phenomenon of ultimate interest to us — depression — we will therefore concentrate on animal data here.

As explained in section 2.4.1, the central psychological postulate of LH, both in its animal (Maier and Seligman, 1976; Jackson et al., 1978, 1979; Weiss et al., 1981; Willner, 1985b; Maier, 1989; Maier and Watkins, 2005) and human Roth and Kubal 1975; Blaney 1977; Costello 1978; Depue and Monroe 1978; Miller 1979 version, is that subjects infer their extent of control in one setting and generalise it to others. Indeed, in chapter 4 we present a model that formalises precisely in Bayesian terms what it means to the goal-directed controller to *explicitly* assess and generalise the degree of control the subject has about its environment. However as we discussed in section 1.2, there is an important interaction between habits and goal-directed systems. Habits also show important generalisation and are likely to be influenced strongly by a number of consequences of LH induction, as is also attested by the provenance of the LH paradigm from probes into habitual learning (Overmier and Seligman, 1967; Peterson et al., 1993). In this chapter, in keeping with the later chapters, we examine the extent to which blunting and the habit system can explain the phenomena of LH and the extent to which contributions of the goal-directed system are critical.

Part of the controversy around the importance of an explicit use of a measure of control has centred on alternative, potentially simpler, explanations of the data. In computational terms foremost amongst these is that shocks can lead to deep analgesia (Terman et al., 1984; Kelly, 1986; Maier et al., 1982; Drugan et al., 1985; Maier, 1989). Parts of the escape deficit might well be accounted for if subjects generalise their level of analgesia between scenarios, which is highly feasible given the rich learning repertoire displayed in tasks involving analgesia (Amit and Galina, 1986; Kelly, 1986). Analgesia is of course one aspect of the insensitivity to primary reinforcers for which we reviewed the evidence in chapter 2, and as such represents in and of itself an interesting object to model. But it is particularly attractive in the context of LH, because it provides us with information about the kinds of reward statistics that impel subjects to choose analgesia.

Here, thus, we will explore an internally directed action — such as the release of endorphins — which reduces the impact of negative reinforcers at an evolutionarily fixed cost. The availability of this action will be combined with learning algorithms that make no explicit reference to any notions related to control. Jointly, this will allow us both to assess the impact of such an action on learning and to further specify precisely which aspects of the behaviour might better be explained by a learning system that has access to, and can make use of, explicit information about control, such as the goal-directed system.

The reinforcement-learning models we use are standard, cached models (see section 1.2 and appendix A.3). In cached models, the value of actions is learned by averaging the long-term outcomes over repeated choices of an action (Sutton and Barto, 1998) and (usually) carry

no explicit information about issues like outcome entropy, uncertainty and so on (Baum and Smith, 1997; Sutton and Barto, 1998; Dearden et al., 1998). Cached systems, like habits, are insensitive to sudden changes in the reinforcing value of outcomes (as sudden changes only have a small effect on the long-run average of a value), unlike decisions made based on tree search and to goal-directed actions, which rely on a model of the environment, rather than on extensive experience in it (Daw et al., 2005). We present models of the minimal complexity needed for the experiments and proceed by way of elimination, simply attempting to model, in a sequential and incremental manner, increasingly complex aspects of LH. We will argue that what is left needs the development of yet more powerful computational tools. Throughout, it is the generalisation behaviour which will provide the challenges. In the first two models presented here, the generalisation behaviour will be built into the state-space description of the models. In the third experiment, a more explicit approach to generalisation will be taken.

Three experiments are modelled here. In the first experiment we introduce our formulation of blunting and show that it replicates some basic data on analgesia in LH. In the second and third experiments, we consider two aspects of generalisation at the heart of LH: valence generalisation in experiment two, and generalisation across environments (or trans-situationality; Maier and Watkins 2005) in experiment three. By showing that both of these latter experiments are replicable with our formulation of blunting, and without access to any explicit reference to control, we argue that the notion of control may be sufficient, but is not necessary to explain a large number of experiments on learned helplessness. However, we were unable to replicate the generalisation of learned helplessness across reinforcer valence using just the notions of blunting and shocks, and this will form one cornerstone of our subsequent treatment of control in chapter 4.

3.2 SHOCK SIZE IN LEARNED HELPLESSNESS

The main inspiration for the models that follow is displayed in figure 3.1. Figure 3.1A (Maier, 1989) shows that exposure to painful stimuli (here 5 electric shocks) can by itself *in*crease the threshold for subsequently applied painful stimuli. The effect of reinforcers is modulated by the history of experienced reinforcers, showing that reinforcers themselves can interact. Figure 3.1B (Jackson et al., 1978) shows the corresponding effect in a LH setting. Here, rats are placed in restraining tubes and either given (inescapable) shocks or no shocks before being given escape training. As can be seen, the rats that have been shocked fail to escape when shocks of 0.6 mA are delivered to the floor of the shuttle box, while unshocked rats escape readily. However, as the shock size is increased, even the rats that were shocked acquire the escape response. It seems that this is precisely the same effect as observed on the left: a response (escape) is now only emitted for higher shock levels.

3.2.1 MODEL DEFINITION

The specific model we use is highly abstract and minimal and is shown in figure 3.2. As usual in reinforcement learning, the task is represented by a set of *states*. For example, in this case we only have two states, characterised by whether the shock (of size S_0) is being delivered (`shock`)

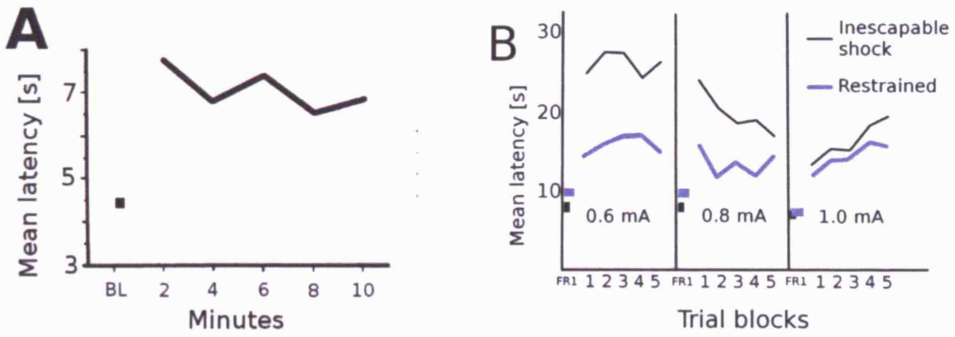
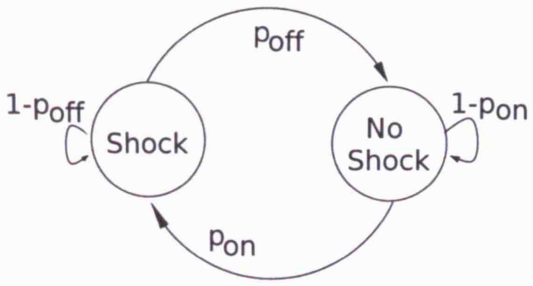


FIGURE 3.1: **A**: Pain-induced analgesia. The figure shows the tail-flick latency of a rat before (BL) and after 5 electric tail-shocks of 5 seconds duration and 1mA strength. Clearly, the pain threshold is elevated for a prolonged period of time. Adapted from figure 3 in Maier (1989). **B**: Effect of shock size on escape behaviour. The graphs show data from three escape experiments. The black (top) lines show escape latencies for rats that have been exposed to inescapable shocks in a restraint tube, the blue lines are for rats that were only restrained in the same tubes but not shocked. Throughout, the unshocked rats escape rapidly. At low levels of shock, the rats exposed to IS do not escape (there was a time-out of 30s), or escape very slowly. As the shock size is increased, their escape latency approaches that of the unshocked rats. Adapted from Jackson et al. (1978).



Actions:

- do nothing — no cost
- escape — cost $C(e)$ affects p_{on}, p_{off}
- blunt — cost $C(b)$ affects reinforcements

FIGURE 3.2: Model of the basic LH paradigm. There are two states: `shock` and `no shock`, and transitions are stochastic, with probability p_{on} and p_{off} from and to the shock state respectively. p_{off} is increased by the escape action, while the blunt action decreases the size of the shock. Only the do nothing action, which has no effects, incurs no costs.

or not (no shock). Shock onset happens when agents move from the no shock state to the shock state. In the model, this happens with a constant probability p_{on} per unit (of discrete) time, meaning that shocks come on at random intervals. Shocks last until the agent returns to the no shock state, and this again occurs with some smaller probability p_{off} per unit time. These transition probabilities can now be affected by actions. In the model, an action is chosen at every time-step, and only has effect for that time-step.

The relevant outcome of all escape actions, be it the turning of a wheel, the pressing of a lever or the shuttling to the other side of a shuttle box is that it shortens the length of a shock. If the shock is escapable, an escape action can be simply implemented as an increase in the probability p_{off}, i.e. an increase in the rate with which agents move back from the shock to the no shock state. If the shock is not escapable, even performing an escape action will not increase the rate at which the shock turns off. On the other hand, a waiting action, or doing nothing, leaves the transition probabilities unchanged, whether the shock is escapable or not. Thus, the actions here really represent *classes* of actions (or mega-actions as they might be used in hierarchical reinforcement learning; Dietterich 2000), and this representational choice subsumes some important issues of generalisation to which we will return later. Escaping is physically taxing, especially escaping fast. To reflect this fact, each escape action incurs a cost C_e; in fact we allow actions to be performed at (three) different levels of effort x, with larger changes in p_{off} and larger costs $C_e(x)$ for escapes with more effort x. This cost is zero for the waiting action.

Finally, there is the blunting action mentioned above. Unlike traditional actions in reinforcement learning (RL) settings, this action does not take its effect via the transition probabilities, but via a change to the subjective impact of the shock obtained when in the shock state. If this were the only effect it had, it would of course always be chosen (the aim is of course to choose the actions with the smallest long-term cost, or equivalently the largest long-term rewards) at the expense of other actions, or in combination with other actions if agents are allowed to choose both simultaneously. If there were no cost to it, organisms should always choose to live in the painless bliss of endorphin, despite the fact that this rapidly leads to a deterioration of an organism's physical abilities and death. To prevent this outcome, a cost C_b is added each time the blunting action is chosen. This cost is assumed to be fixed and have been learned on an evolutionary time-scale to represent the resultant long-term severe adversity of life without pain perception (Boureau, 2005). Again, agents can choose to blunt more or less. The larger the blunting "effort" x, the smaller the perceived shocks, but the larger the cost $C_b(x)$.

LH rests on generalisation (Seligman and Maier, 1967; Maier and Seligman, 1976; Maier and Watkins, 2005): subjects learn in one environment and then use aspects of the acquired knowledge in the new environment. In animal experiments, rats learn about one type of escape (e.g. turning a wheel) and generalise this to other escape actions (e.g. shuttling). Here, we take the most drastic approximation to this, and assume that animals learn about the value of escape actions in general, but have this information be applied automatically to whatever escape action is appropriate in a particular environment. In that case, the generalisation simply involves a reversal of contingencies: rats that were in the uncontrollable scenario suddenly have control. In the experiment of figure 3.1B, animals were given either no shock or inescapable shock. The use of no shock animals stems from the earlier finding that animals which are given

escapable shock do not differ from animals that are not exposed to any shocks (Jackson et al. 1978, though in other scenarios there can be mastery; Peterson et al. 1993). We will relax this approach to generalisation in a subsequent part of this chapter, and then further in the next chapter.

3.2.2 LEARNING

After training, animals know what action to take in which state. A mapping from states to actions is a policy. Here, we learn stochastic policies based on the quality values $Q(s,a)$ of actions in states. The value $Q(s,a)$ of taking an action a in state s is equal to the sum of all expected future reinforcements when taking that action a in that state s (see appendix A). Learning *cached* values $Q(s,a)$ just involves book-keeping: all reinforcements experienced after each occurrence of a state-action pair are simply averaged (Watkins and Dayan, 1992; Sutton and Barto, 1998). However, if, as in our case, there is no obvious delimitation of trials (agents simply move between the two states and take actions in a continuous manner; there is no discrete onset or end to a trial), these values can become arbitrarily large, as the sums run over long times. One option is then to use average-reward cached methods (Mahadevan, 1996). Here, the expected total reward averaged across actions is subtracted from all actions, and actions are chosen according to their *advantage* dQ. An iterative update that achieves this is called average-reward SARSA. Specifically, we write the advantages $dQ_t(s_t,a_t)$ at time t as

$$dQ_t(s_t,a_t) \leftarrow dQ_t(s_t,a_t) + \varepsilon\delta \tag{3.1}$$

$$\text{where} \quad \delta = dQ_t(s_{t+1},a_{t+1}) - dQ_t(s_t,a_t) + r_t - \rho_t \tag{3.2}$$

$$\text{and} \quad \rho_{t+1} = \rho_t + \epsilon(r_t - \rho_t) \tag{3.3}$$

where s_t is the state visited, a_t the action taken, and r_t the reinforcement obtained at time t. The parameters ϵ and ε are learning rates, with $\epsilon < \varepsilon$. Thus the advantage of an action is difference between the expected total future reward for that action and the average expected total future reward over all actions, which is represented by ρ_t. The sum of ρ_t and dQ_t is an approximation to the true value Q. The advantage dQ of actions is updated by δ, the prediction error. Because blunting represents a nonlinear interaction between reinforcers, the Bellmann equations (see appendix A.3) can no longer be solved analytically, and we have to resort to simulations.

Given dQ values, we can now write down the policy. The probability of choosing action a in state s is:

$$p(a|s) = \frac{\exp(\beta\, dQ(s,a))}{\sum_{a'} \exp(\beta\, dQ(s,a'))} \tag{3.4}$$

where β is a parameter that renders action choice more or less stochastic. Thus, actions compete for expression via their Q-values. As $\beta \to \infty$, the action with the maximal dQ-value is deterministically chosen, while as $\beta \to 0$, actions are chosen randomly irrespective of their dQ values.

Thus, cached, habitual learning proceeds as follows: the agent initially just chooses actions at random (all the Q values are equal and zero), but over time some actions in some states come

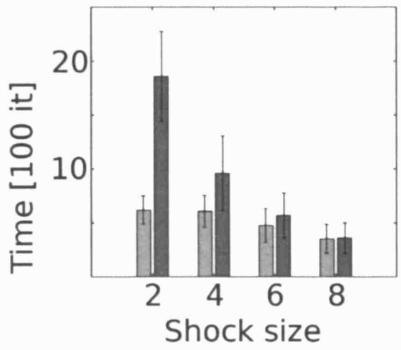


FIGURE 3.3: Reversal latency as a function of shock size after inescapable or escapable shock. The light bars represent escape latencies after escapable shocks, the dark bars after inescapable shocks. Shock size only has strong influence on the acquisition of an escape after exposure to inescapable shocks.

to be associated with more future rewards, and are therefore chosen more often. In psychological terms, a stimulus (state)-response mapping is acquired — alternatively also known as a habit.

3.2.3 RESULTS

Figure 3.3 shows a replication of the basic effect in figure 3.2B. In the model, Q values for the actions blunt, escape and wait are first learned for a shock size $S_0 = 2$. The escapable and inescapable configurations differ only in whether escaping alters p_{off}. A reversal then occurs: escape actions either become effective or lose effectiveness. Furthermore, the shock size changes. Consider first the switch from inescapable shocks of size $S_0 = 2$ to escapable shocks of equal or greater magnitude, displayed by the dark bars in figure 3.3. The bars represent the time until agents' Q values for escaping exceeds that of blunting or waiting. The larger the shock, the shorter this time. On the other hand, the light bars in figure 3.3 represent the time it takes for agents who come from the escapable situation to learn that the escape action is no more functional, and that blunting is the optimal option. We see that this delay is affected to a much lesser degree by the shock size.

Figure 3.4 explores the results in more detail. It shows the time-course of the Q values for all actions before and after the reversal, for the case where the shock size is held constant at $S_0 = 2$. The insets of figures 3.4A and B show that, after prolonged training, it is best not to do anything in the `no shock` state, whether shocks are escapable or not. This is correct: when the shock is off, there is nothing to gain from the expensive escape or blunting actions. The insets of figures 3.4C shows that after the initial training in the escapable scenario, the most vigorous escape action is chosen, while a long time after the reversal, the most severe blunting is chosen. These are again the correct choices — when the escape does not shorten the shocks by increasing p_{off}, the most beneficial action is the blunting action, which despite incurring some cost $C_b(x)$, alleviates the strength of the shocks. However, when escape actions do increase p_{off} by a sufficient amount, they are the optimal choice. Similarly, the insets of figure 3.4D show that

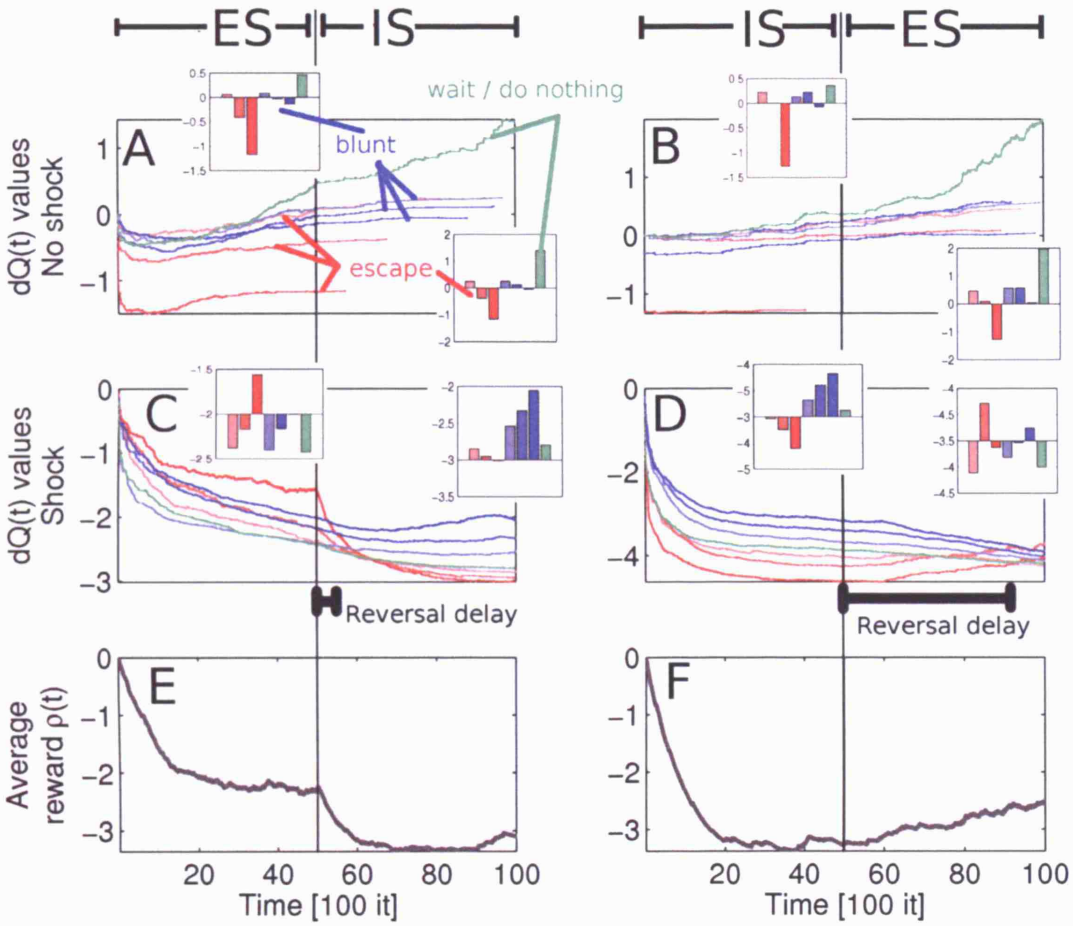


FIGURE 3.4: Model advantage values dQ of all seven actions for each state over time (in 100 iterations). The reversal from ES to IS or vice versa takes place at the dark vertical bars. The seven actions available are three escape actions (red), three blunt actions (blue) and one waiting action (green). Escaping and blunting actions are available at three levels of effort $x = \{1, 2, 3\}$, represented by increasingly dark colours. The insets of each panel show the dQ values at the reversal and at the end of learning (note the varying baselines). Left column: Reversal from escapable to inescapable shock; Right column: reversal from inescapable to escapable shock. **A,B**: values for the `no shock` state. Agents learn to do nothing in this state, and keep doing so after the reversal. **C**: Agents learn to choose the maximum effort escape action when the shock is on. After the reversal, the value of this action drops rapidly, resulting in a short reversal delay before agents choose to blunt. **D**: Agents learn to blunt maximally during IS. After the reversal, they continue blunting for a long time. The black bars below panels C and D indicate the reversal times, means of which over 100 learning trials are displayed in figure 3.3. **E,F**: Average rewards $\rho(t)$ for the two experiments. Only when the switch is from escapable to inescapable shock (panel E) is there a noticeable change in the average reward.

Parameter	Value
Shock size S_0	2
Blunted shock size with effort x	$S_0(1-0.2x)$
Cost of waiting / doing nothing w	0
Cost $C_b(x)$ of blunting with effort x	$0.2x$
Cost $C_e(x)$ of escaping with effort x	$0.4x$
Shock onset probability p_{on}	0.2
Blunting and do nothing p_{off}	0.02
Escape p_{off} with effort x	$0.1+0.2(x-1)$
β	4
ε	0.1
ϵ	0.01

TABLE 3.1: Parameter values for results of section 3.2.

after initial training in the inescapable scenario, the most severe blunt action is favoured, but that prolonged training in the escapable scenario after the reversal results in correct acquisition of an escape action.

Figure 3.4A and B themselves show that nothing major happens to the dQ values of the actions in the no shock state at the time of the reversal — learning seems to continue along its previous trajectory. This is very different for the shock state. Figure 3.4C shows that the best action (the red line representing escape with maximal effort $x = 3$) very rapidly loses value after the reversal: the costly escape action, which was worth it when it did reduce the time spent exposed to shock, is very rapidly found to not be effective any more, and its value decreases beyond that of the blunting actions. Figure 3.4D on the other hand shows that the dynamics of the dQ values after the reversal are very different when an action of the type blunt is chosen frequently, as it is after training in the inescapable scenario. Now, the altered contingencies at $t = 50$ are not so apparent: The costly escape actions are rarely explored, due to their low dQ values, and when they are chosen, their advantage over the blunting action is not that great. Finally, figures 3.4E and F show the dynamics of the average values. Again, transfer from the escapable to the inescapable scenario results in drastic changes in the reward statistics, whereas the opposite switch only produces a very small change, which has to accumulate for a while before behavioural alteration is actually observed. The reversal latencies in figure 3.3 are the times from the reversal until the optimal action in the new environment achieves maximal dQ value and are indicated by black bars below figures 3.3C and D.

For completeness, the parameter values for the results presented here are in table 3.1. The crucial values are the shock size S_0, which here is arbitrarily chosen as 2, and the p_{on}, which determines how likely shocks are. The results presented here are robust to changes that maintain the approximate relative sizes of shocks and costs.

Thus, a change in the world from controllable to uncontrollable shock is much more evident than the opposite, once policies have been acquired. Use of the actions that decrease the sizes of reinforcers, such as pain-induced analgesia, have profound effects on the speed with which new policies can be acquired and lead to hysteresis in policy space. In a normative setting one can think of taking recourse to actions that diminish the size of reinforcers as an analogy to making the assumption that the best policy has been achieved and is unlikely to change soon.

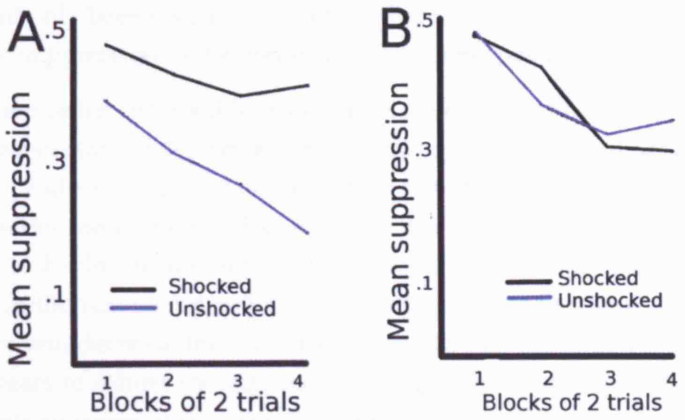


FIGURE 3.5: The effect of inescapable shock on contingency judgement. Rats had acquired an appetitive lever-press response and then been either restrained or restrained and shocked (inescapably). The plots show the ratio of responding during a CS which is associated with shock. **A**: Suppression ratio, over time, of rats that are exposed to the contingent, escapable shock condition. Here, rats can avoid the shock if they do not press the appetitive lever during the CS. Smaller suppression ratio means rats suppress more. **B**: Suppression ratio of rats exposed to the non-contingent, inescapable shock condition. Here, shocks are delivered during the CS independently of the rats responses. The rats not exposed to shock show a faster response suppression in the contingent condition than the shocked rats, while the two groups don't differ in the non-contingent condition. Adapted from Jackson et al. (1978).

3.3 CONDITIONED SUPPRESSION

The previous section illustrated the basic issues surrounding the availability of actions that alter the effects of subsequent reinforcers, such as blunting or analgesia, and showed how such an internally directed action can account for the shock size of the escape deficit after IS. Here we proceed to model an ingenious experiment (Jackson et al. 1978, experiment 4) which argues more strongly and directly for a perception of control as being the relevant variable necessary to explain the LH effects.

Jackson et al. (1978)'s experiment proceeded in three steps. Rats were first trained on an appetitive lever-press schedule for food. After rats had acquired the response, they were given either escapable or inescapable shock in a different environment. All rats were then returned to the lever-press cage. Each of the two groups was then subdivided into two further groups. For one half of the rats, a CS would come on (randomly) and predict two inescapable shocks. For the other half of the rats, the CS would indicate that two shocks would be delivered contingent on lever-pressing. Thus, for the second set of the rats the shocks are entirely avoidable. Figure 3.5A shows the suppression ratio (the ratio of lever presses before the last part of the experiment and during the CS) when rats were exposed to the shocks contingent upon lever presses. The rats that have had previous exposure with inescapable shocks suppress less rapidly than

the rats which had only been restrained. On the other hand, exposure to previous shocks does not alter response suppression in the non-contingent, inescapable scenario.

These results are rather remarkable. First, this is a nice control for a long-standing objection to learned helplessness (and indeed an alternative theory of depression), which is that exposure to shocks simply produces a general depression of activity. Here, rats that have been exposed to inescapable shocks respond more than the unshocked rats. Second, the effect in figure 3.5B is usually termed a Pavlovian-Instrumental Transfer (PIT). A priori there is no reason for the rats to suppress in the non-contingent case, as decreasing the response will not decrease the punishments, but will decrease the rate of rewards. In PIT, the negative Pavlovian value of the CS itself appears to inhibit the actions. A straightforward interpretation of the fact that both groups of rats suppress at the same rate is that they must have come to assign the same Pavlovian value to the CS, and thus be equally sensitive to the reinforcer.

3.3.1 MODEL

To model this experiment, some of the stringent simplifications introduced in the previous model have to be relaxed. Firstly, the experiment is more complex, and needs a larger number of states. Then, consider again the previous model: the generalisation required by the experimental manipulation was represented purely by a switch of outcome contingencies for the escape class of actions. All action classes kept their state-action $\mathcal{Q}(s,a)$ values across the reversal. The formulation in terms of action classes implemented the generalisation from one kind of escape action to another. We find that the generalisation issues here can be accommodated by a representation which allows the learning of *one* $\mathcal{Q}$ value for blunting actions across states, i.e. actions a like pressing a lever, escaping or waiting will have standard $\mathcal{Q}(s,a)$ values which are bound to a particular state, but the blunting action b has a state-independent value $\mathcal{Q}(b)$. Thus, animals in this model will be able to both blunt and do some other action like pressing a lever or escaping. Learning to blunt in one part of the overall experiment means that the animals will start by blunting in subsequent parts of the experiment (though they can still learn not to blunt).

The expanded model is shown in figure 3.6A. There are now five states, corresponding to the five informative states the animal can be in in the three parts of the experiment. For the first, appetitive learning, part of the experiment, there is just one state (`l` for lever, bold in figure 3.6A). The rat can blunt, press the lever, which leads to rewards with some probability, or it can wait. The most rewarding action is to press the lever. Lever pressing and waiting have values that are bound to the state `l`, but blunting has a value which is independent of the state.

In the second part of the experiment, rats are constrained to the states `n`, which is simply an empty cage, or `s`, which is an empty cage with shocks being delivered (bold states in figure 3.6B). Transitions between states are stochastic, just as in the previous section. For the rats in the escapable condition, choosing to escape increases the probability of moving from `s` back to `n`, while for rats in the inescapable condition it has no such effect. Rats in the inescapable condition will come to choose the action combination of blunting and waiting. There is no lever and no lever pressing in these two states. In the third part of the experiment (figure 3.6C), rats are returned to the cage with the lever, but an additional two states are available: `l, c` when

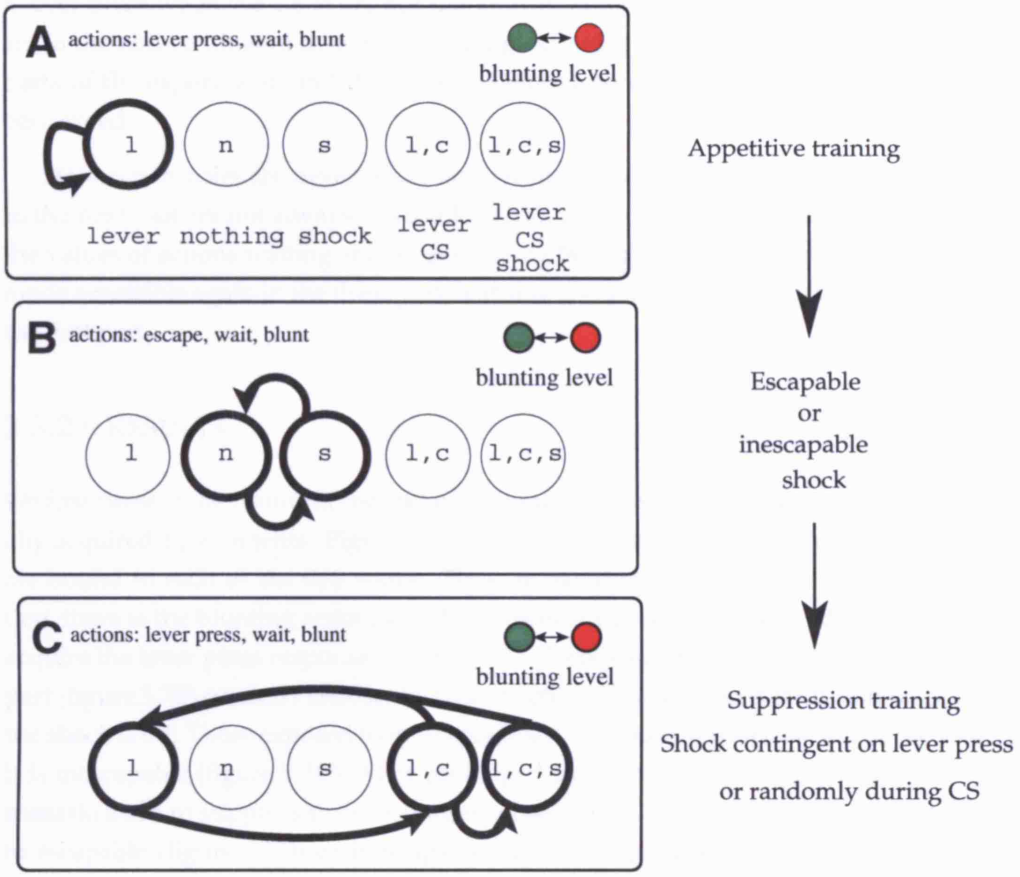


FIGURE 3.6: Model of the Jackson et al. (1978) experiment four. The states available in each part of the experiment are the circles in bold. Arrows represent all possible transitions (due to all available actions) between states in the respective part of the experiment. The value of blunting, represented by the green and red circles in the top right corner, are generalised across all parts of the experiments. They are mainly being used in parts two and three of the experiment. **A**: Appetitive lever press training. Only one state is available, and it is characterised by the presence of a lever but nothing else. **B**: IS/ES exposure. Two states are available, characterised by the presence of a shock or nothing at all. Rats in the inescapable scenario learn to blunt here. **C**: Suppression training. The state with only a lever is again accessible, as are one state with an additional CS, and one with lever, CS and shock. The values for blunting are generalised from the previous parts of the experiment.

both lever and CS are present, and `l, c, s` when lever, CS and shock are present. Rats can lever press or wait and blunt. Transitions from `l` to `l, c` are stochastic and independent of actions, with a fixed probability per unit time. Throughout, lever presses yield rewards. In the contingent scenario, shocks will come on (i.e. rats will move from `l, c` to `l, c, s`) only if they perform a lever press when the CS is on, but this transition will occur after a random delay if the rats are in the non-contingent scenario. Escaping, blunting and pressing the lever incur costs in all parts of the experiment, and these are as before a function of the effort with which they are performed.

State-action pairs are never reset, transfer their Q values from one stage of the experiment to the next, but are not always accessible. Thus, in the first part, only state `l` is accessible, and the values of actions waiting and lever pressing (and blunting) are acquired. When the state is made accessible again in the third part, learning continues starting from the values acquired in the first part.

3.3.2 Results

We first present the results at the end of training to show that the correct responses are eventually acquired by all agents. Figure 3.7 shows choice probabilities for the actions whose values are bound to each of the five states. Thus, in each state, two actions are available. In addition, there is the blunting action, which is however not shown here. In the first part, all agents acquire the lever press response (figure 3.7A). Those exposed to escapable shock in the second part (figure 3.7B) correctly choose the escape action when the shock is on, and do nothing when the shock is off. Those exposed to inescapable shock choose to wait (and blunt, see below) when it is inescapable (figure 3.7C). After prolonged suppression training, agents in the contingent scenario learn to suppress their responses when the CS is on, whether they have been exposed to escapable (figure 3.7D) or inescapable (figure 3.7F) shock. Agents in the non-contingent scenario continue preferring to press the lever over inaction, irrespective of the previous treatment (figure 3.7E,G). However, the presence of the CS does decrease their propensity to press the lever, which replicates the PIT seen in figure 3.5B.

Figure 3.8 shows the Q values for blunting and not blunting. In the escapable scenario, it is advantageous not to blunt, while it is advantageous to blunt when shocks are not escapable. Note that the average value is much lower when shocks are inescapable. These are the Q values that are generalised to the next part.

While figure 3.7 showed that after extended training the impact of blunting is overridden and all agents acquire the same response patterns, blunting does lead to differential delays in the contingent response suppression without affecting the non-contingent time courses — which is in essence the result of Jackson et al. (1978)'s experiment four. Figure 3.9A-D shows Q value time courses for a total of 100 agents. Comparison of the time courses for the two actions in figure 3.9A and C shows that, the *difference* between the Q values of the two actions grows much more slowly after the IS than after the ES. Blunting does not only affect the active response, but also decreases the "relief" for the passive response. Figure 3.9B and D show that in the non-contingent scenario the lever press actions lose value at the same, slow rate whether they were preceded by IS or ES.

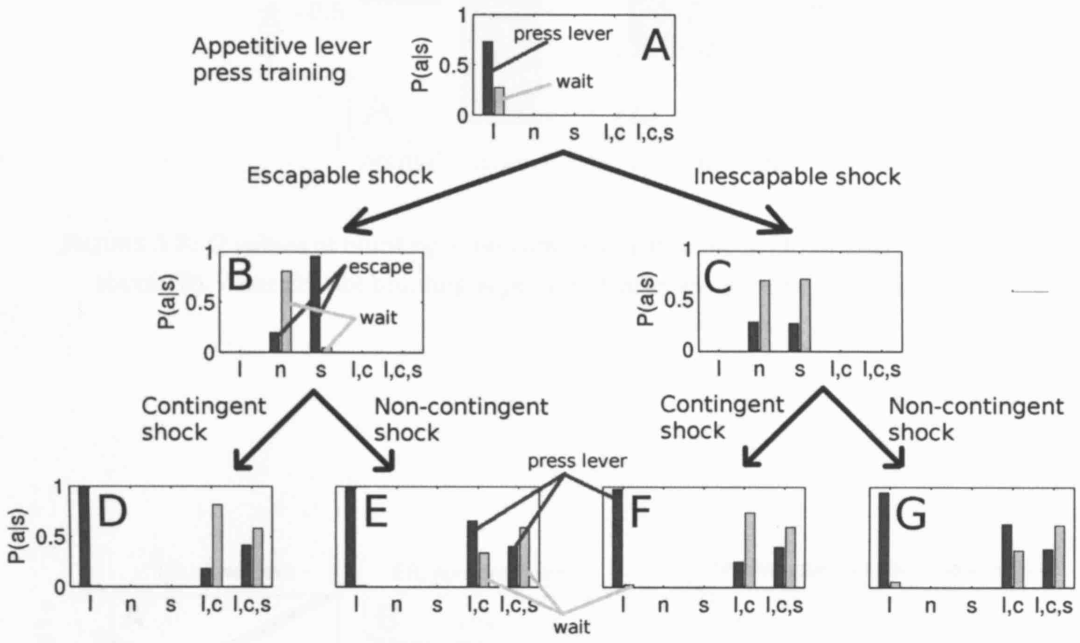


FIGURE 3.7: State-action probabilities at the end of each of the three stages for all experimental groups. Each box shows the probability of choosing the active action (escape or press lever, dark bar), or the passive action (do nothing, light bar) for each of the states that are available in the particular part of the experiment (l=`lever`, n=`nothing`, s=`shock`, l,c=`lever + CS`, l,c,s=`lever+ shock + CS`). The values for blunting are shown separately in figure 3.8. **A**: all agents acquire the lever press response; **B**: For escapable shocks, agents learn to do nothing in state `n`, and to escape when the shock is on; **C**: agents learn to just wait in both states when shocks are inescapable. **D,F**: Agents in the contingent shock scenarios choose to do nothing rather than pressing the lever when the CS is on and when the shock is on (states `l,c` and `l,c,s`). **E,G**: Agents keep pressing the lever when the CS is present, but less so than when it is absent. This reduction represents the PIT.

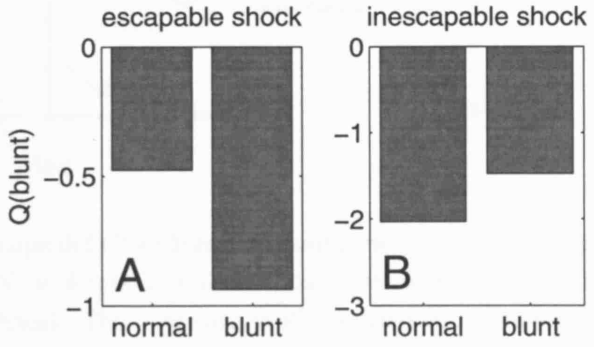


FIGURE 3.8: Q values of blunting action after escapable shock (**A**) or after inescapable shock (**B**). After ES, not blunting is preferred, after IS, blunting is preferred.

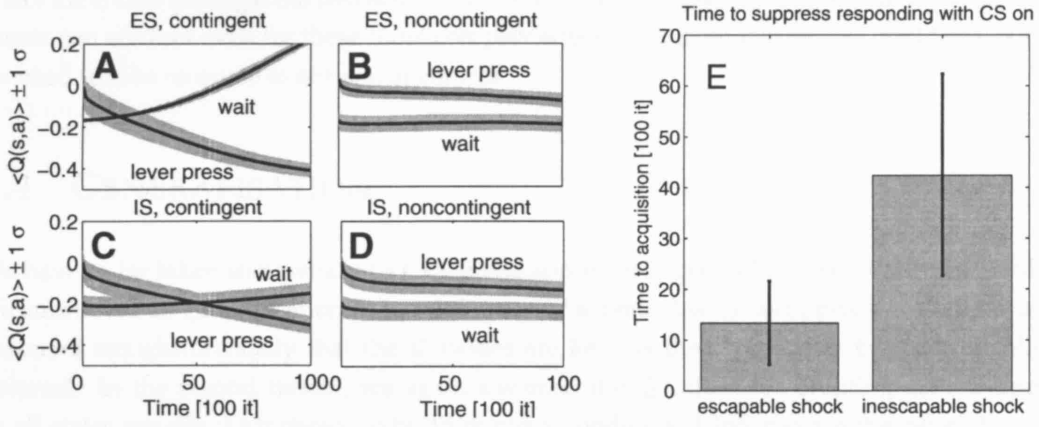


FIGURE 3.9: Blunting delays response suppression in the contingent but not in the non-contingent scenario. Temporal evolution of Q value for pressing lever (dashed) and doing nothing (solid) **A**: after escapable shock in the contingent suppression scenario; **B**: after escapable shock in the non-contingent scenario; **C**: after inescapable shock in the contingent scenario and **D**: after inescapable shock in the non-contingent scenario. The grey zones are $\pm$ 1 standard deviation over 100 subjects. **E**: Time until the response preferences switch in the contingent scenario after escapable or inescapable shock.

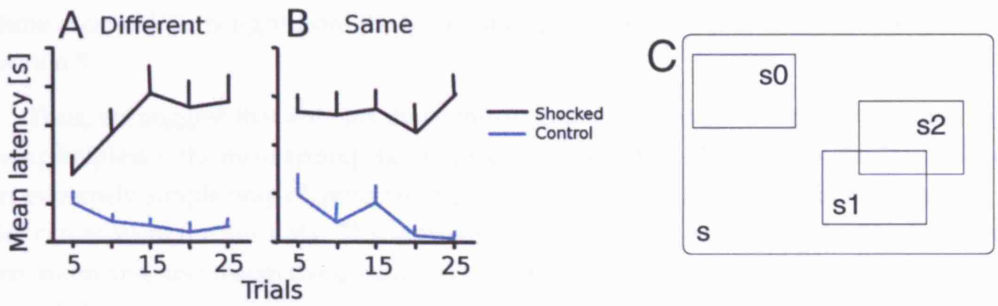


FIGURE 3.10: Escape deficit with and without generalisation. The figures show escape latencies of $N = 8$ rats in a shuttlebox. Rats were given IS (black lines) or no shock (blue lines). The exposure to IS occurred in a different environment (**A**), or in the shuttlebox itself (**B**). **C**: generalisation schema: within some space s that defines distances between environments, it may be beneficial to rely more on close environments than on distant ones. A and B adapted from Maier and Watkins (2005).

The effect of blunting is now easily understood. In the contingent scenario, the lion's share of the difference in value between pressing and not pressing the lever when the CS is on is carried by the shocks. Blunting will diminish this difference. On the other hand, in the non-contingent scenario, the shock does not differentiate between the actions. Blunting will not affect the choice amongst the two actions. Thus, we find that alteration of sensitivity to punishments can account even for these more complex effects of inescapable shocks, and there is still no need to take recourse to notions of control.

3.4 GENERALISATION

We have so far taken somewhat of a Cinderella approach to generalisation. In our first model, we subsumed all generalisation in our definition of actions classes (as opposed to actions) and assumed straightforwardly that the Q values are kept as they were after the controllability reversal. In the second model, we again assumed the Q values for blunting actions apply to all states equally. Our choice to build minimal models and incorporate the generalisation in the way the state-space was built was made in order to isolate the effects of blunting and controllability.

However, the models are lacking in that they were not designed to account for more specific effects of generalisation that are also central to learned helplessness. Seligman (1975) gives a host of very general human examples, and Maier and Watkins (2005) give some specific examples in animals which will be the aim of our third model. Consider figure 3.10, which shows shuttle box escape latencies of rats that have received inescapable shock. Figure 3.10A shows escape latencies for rats that were given shocks in a different environment, and B shows the latencies for the rats that were given inescapable shock within the shuttle box in which they were then tested. Rats that were tested in a different environment initially escape *quickly*, and slowly, over about 10 trials, *develop* their escape deficit. Rats that were shocked within the shuttle box

show escape deficits right from the beginning. It is only this latter effect that we modelled in section 3.2.

Thus, we suggest that animals that come to a new environment first have to infer whether being helpless is the most appropriate course of action to take. In the following, we will see that an extremely simple way of generalising between environments on the basis of reward statistics can account for this data. This consists in comparing distributions of rewards in different environments, and weigh the $\mathcal{Q}$ functions of the various environments accordingly to produce a prediction in a new environment. Consider figure 3.10C: $s0$, $s1$ and $s2$ are environments with reward characteristics that determine their position with the large space s of all possible such environments. Let us assume, an agent has experience of environments $s0$ and $s2$. When faced with environment $s1$, about which it has no knowledge, it should rely more on the closer environment $s2$ than on $s0$. In the following we formalise just this intuition, and use the reward statistics of environments to position environments within this space and to define distances between them.

3.4.1 METHODS

Specifically, we assume that $\mathcal{Q}^i$ values are known for some set of environments $i = \{1, 2 \cdots E\}$, and that a few observations have been made in the present environment e leading to a rough, small-sample estimate of $\mathcal{Q}^e$. Let us assume that we generalise between classes of actions, rather than between particular actions. When we say "escape" we will have in mind all actions that would lead to escape, and we will relate the $\mathcal{Q}$ values of action classes a directly to each other. When acting in environment e, which is rather unknown to the agent, we will assume the agent constructs an *effective* $\mathcal{Q}$ value by which to choose amongst classes of actions. We will represent the effective $\mathcal{Q}$ value for the present environment as a weighted mixture of the $\{\mathcal{Q}^i\}_{i=1}^E$ values of action classes in other environments, and the rough $\mathcal{Q}^e$ value inferred about that class of actions from the limited information gained in the present environment. As more information is gained, we of course want to rely more on the information from the present environment than on information from other environments.

Let us weigh the contribution of environment i's $\mathcal{Q}$ value to the effective $\mathcal{Q}$ value in the present environment e by the fractional Kullback-Leibler divergence (D_{KL}) between the present distribution of rewards and the reward distribution in the environment i, i.e. let w_i be

$$w_i^* = D_{KL}(P(r|e)||P(r|i)) \tag{3.5}$$

$$w_i = \frac{w_i^*}{\sum_j w_j^*} \tag{3.6}$$

where $P(r|e)$ is the reward distribution in environment e and $P(r|i)$ is that of environment i. Thus the contribution from all other environments to the efficient $\mathcal{Q}$ value is

$$q(a) = \sum_{i \neq e} w_i \mathcal{Q}^i(a)$$

although we let $\mathcal{Q}$ stand for $d\mathcal{Q}$ (the values are still acquired by average-reward learning) for notational clarity. Furthermore, we define $\lambda(t)$ as an increasing function of the number of ob-

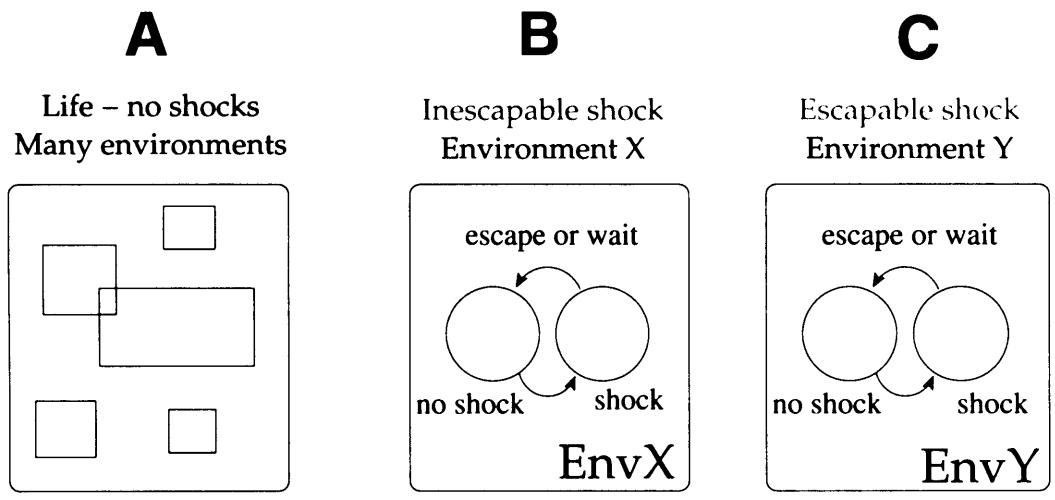


FIGURE 3.11: Generalisation. **A**: Agents first gain experience of many environments (here around 10), with various reward distributions, and of varying similarity. None of these include shocks (severe punishments). Agents learn to be active in all of these environments, and in none do they learn to blunt. **B**: Exposure to environment X with large, inescapable shocks. Agents learn to blunt. **C**: Exposure to escapable shocks in environment Y. This environment shares equally many features with the environments from A as environment X, but only shares the occurrence of shocks with X. Throughout there are thee kinds of actions, which share $\mathcal{Q}$ values: do nothing, be active (escape) and blunt.

servations made in the present environment $\mathcal{Q}$. This will take into account that as more observations are made in environment e, more weight should be given to $\mathcal{Q}^e$ as opposed to the other $\{\mathcal{Q}^i\}_{i\neq e}$. The effective $\mathcal{Q}^e_{eff}$ value in environment e is then

$$Q^e_{eff}(s,a) = \lambda(t)Q^e(s,a) + (1-\lambda(t))q(a) \tag{3.7}$$

Of course, it is possible to extend the formulation in equation 3.5 and include other features:

$$w_i = \psi\left(||s-\sigma||, D_{KL}(P(r|e)||P(r|i))\right) \tag{3.8}$$

where s might be the feature vector of environment s_i and σ that of environment e.

We again attempt to use as simple a state specification as possible. Figure 3.11 shows our choice. First, agents are exposed to a series of environments, in which they experience positive and negative reinforcements drawn from a broad Gaussian distribution. They are then exposed to environment X in which they are given inescapable shock and learn to assign large value to blunting. The reward distribution for this environment is highly peaked around a very negative value. Finally, they are brought to environment Y in which the shocks are escapable, but the reward distribution still has a peak at large negative values and therefore matches that of environment X best.

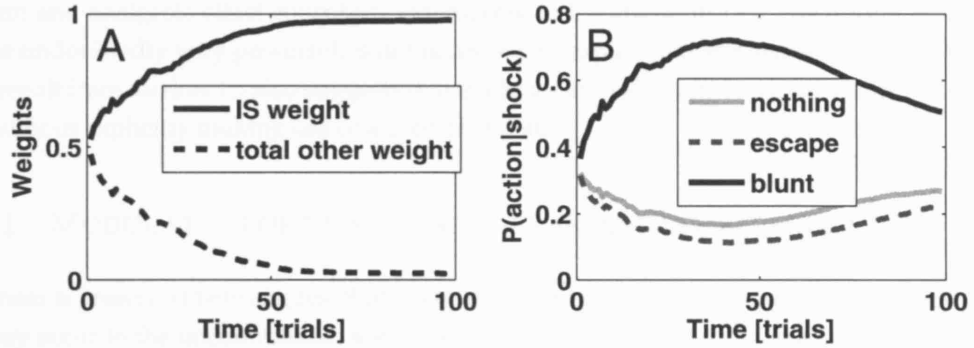


FIGURE 3.12: Temporal evolution of learned helplessness in the model. **A**: Temporal evolution of the weights given to the $\mathcal{Q}$ value acquired in the inescapable shock environment, and to all the other environments. Over time, due to the high frequency of large punishments, only the IS environment is given non-zero weight. **B**: Time course of action choice. Initially all three actions are chosen with equal frequency, but the large weight given to the IS $\mathcal{Q}$ value rapidly produces a strong preference for blunting. The decrease in the preference for blunting despite the maintained large weight for the IS $\mathcal{Q}$ values is due to $\lambda(t)$, which weighs evidence from the current environment more than evidence from others.

3.4.2 RESULTS

Figure 3.12 shows the temporal evolution of the weights and the resulting action choices. In figure 3.12A, we see that, as experience in the escapable shock environment proceeds, the distribution of reinforcements comes very close to that of the inescapable shock environment, and comes to be, at least relatively speaking, further and further away from all the other environments. Panel B shows the resulting action choice. Initially, blunting, escape and inactivity are chosen equally frequently, but as the large punishments are rare in all but the IS environment, the agent rapidly starts relying on the IS $\mathcal{Q}$ values and chooses to blunt. It is only over time, when enough experience is collected, that escape finally starts taking place. The functional form of λ completely specifies how the preference for blunting disappears and is eventually replaced by escape, and is left unspecified here as there are no clear experimental data on this part.

3.5 DISCUSSION

Highly simplified reinforcement-learning models were applied to learned helplessness. There were two aims: to explore the consequences of the availability of blunting as an action and to see whether this, in combination with models that have no explicit knowledge of control can account for the generalisation effects ascribed to control.

We found that the availability of a blunting action has long-term hysteretic consequences for policy selection and that very simple formulations can reproduce three paradigmatic aspects of learned helplessness data. We conclude from this that both the habitual / cached learning

system and analgesic effect contribute extensively to the effects in LH. The notion of control, while undoubtedly very powerful, is not necessary to explain *aversive* learning deficits after IS. The result from section 3.3 also suggests that evidence involving conflicts can also be accounted for without explicitly making use of a control statistic.

3.5.1 MODULATION OF PAIN SENSITIVITY IN ANIMAL EXPERIMENTS

The results presented here suggest that analgesia (more generally referred to as blunting) should mainly occur in the uncontrollable scenarios. Jackson et al. (1979) provide evidence for this. 80 IS foot shocks produce a short-lasting analgesia which is not evident after 24 hours, whether the shock was escapable or not. But in a classical LH paradigm escape training, the rat is re-exposed to shock, and this re-exposure to mild shock re-induces analgesia, but it does so only in the yoked group. In the master group, mild shocks do not re-induce analgesia, and in fact, ES before IS prevents the analgesia. This is precisely the pattern of results recovered here.

The analgesia is only opioid once there has been enough exposure to IS to produce LH. Prior to that the analgesia is non-opioid and relies on the spinal cord (Maier, 1989). After enough shocks to produce LH, the analgesia is central and can be antagonised by naloxone injection into the encephalon (Peterson et al., 1993, p87). This is not inconsistent with the notion of blunting as an internal action and acquisition of a value for this action may proceed along similar lines as those of other, motoric, actions. In fact, this may also rely on similar neuromodulatory systems: there are strong links between central analgesia and serotonin. For example, Sutton et al. (1997) show that exposure to IS 24 hours previously potentiates the response to low doses of morphine. This effect is abolished by intra-DRN injections of the $5HT_{1A}$ agonist 8-OH-DPAT (which due to activity at inhibitory autoreceptors leads to inhibition of the DRN), and is mimicked by intra-DRN injections of β-carboline. Interestingly, ES before IS also prevents the analgesic effect of morphine (Grau et al., 1981) — an effect that is again antagonised by manipulations of the DRN and serotonin (Bland et al., 2003a,b, 2004).

However, there are many facets of analgesia that are beyond our highly simplified account. Firstly, even early on there are many different analgesic phases with different time-courses and pharmacological sensitivities (Terman et al., 1984; Drugan et al., 1985; Maier, 1989). Pain-induced analgesia may influence not only the acquisition, but also the retention of actions (in particular, avoidance Galina and Amit 1986). Pain-induced hyperalgesia is of course just as prominent as hypoalgesia, if not more (and might be related to effects opposite to learned helplessness; Wortman and Brehm 1975), and can be potently influenced by expectations (Ben Seymour, pers. comm.). The fact that trauma can be followed by both hypo- and hyperalgesia represents, in our view, a major unsolved mystery in pain research, and puts pain into a category apart from the senses and somewhat closer to the intentions. In some ways it may be the only aspect in which aversive and appetitive learning are not mirror images of each other (the reader be reminded of the proverbial soldier in a battle who only notices the missing leg in the field hospital).

Blunting is an action with complex consequences. Figure 3.4 shows that its selection of course leads to insensitivity in environmental contingency changes. Not only does opting to blunt affect future behaviour, learning how much to blunt corresponds, to a certain extent, to

learning how reinforcing reinforcers are. As such it is a clearly ill-posed problem. However, it reflects the ill-posed problem the brain has to solve, as organisms clearly do have access to analgesia. At some level, it is just the age-old, unanswered question about the correct utility function (Kahneman and Tversky, 1979). Here, we have rendered the problem well-posed by making blunting costly. We have not derived this cost from any *a priori* arguments. On evolutionary grounds, one might suggest that this cost should embody the correct trade-off between long-term costs of bodily harm due to absence of pain sensation and the short-term costs incurred by missed rewards lurking behind some pain. This is an important avenue for future research.

3.5.2 SEROTONIN

We have here shown that extensive aspect of LH are reproducible by cached learning from aversive events together with analgesia. Acquisition of the values of actions in the model relies on a predictive error signal (TD error; equation 3.2), the positive part of which is well-known to be carried by the phasic activity of dopaminergic neurones (Montague et al., 1996; Schultz et al., 1997; Schultz, 1998). However, as learning here is mainly from negative reinforcers, it relies mainly on a negative TD error signal, which has been suggested to be carried by serotonin (Daw et al., 2002) based on extensive behavioural and neurobiological evidence (Gray, 1991; Deakin and Graeff, 1991; Graeff et al., 1998; Graeff, 2002; Gray and McNaughton, 2003).

Indeed, serotonin does play a key role in the acquisition of LH and it has been suggested that it is the final common path (Maier and Watkins, 2005) for the expression of the effects of IS. Thus, IS activates serotonergic neurones of the dorsal raphé nucleus (DRN) in a graded manner (Takase et al., 2005) and more so than ES (Takase et al., 2004, 2005); lesion (Maier et al., 1993) and pharmacological (Maier et al., 1995b) inactivations of the DRN or even inactivation of the DRN by descending inhibition from the prefrontal cortex (Amat et al., 2005) all prevent the escape impairment after IS.

However, in our model the acquisition of a successful *proactive escape* response is also dependent on a negative TD signal. One potential resolution might come from the more detailed consideration not only of serotonin's role in the acquisition of Pavlovian values, but also in the direct modification of behaviour. Serotonin is known to impair appetitively maintained actions (Carter and Pycock, 1978; Fletcher, 1996; Kapur and Remington, 1996; Fletcher and Korth, 1999) and promote response suppression in the face of some punishments (Deakin and Graeff, 1991; Gray, 1991; Graeff et al., 1998; Graeff, 2002), but its Pavlovian action arsenal also extends to proactive actions, such as escape and defensive aggression (Blanchard and Blanchard, 1988). We will show in chapter 5 that aspects of this dual role in both evaluation and action selection can have profound effects, which will likely be important to account for the differential sensitivity to 5HT of escape after IS and ES.

3.5.3 GENERALISATION

Our simplified approach to generalisation (relying on distributions of experienced rewards) can account for the rather complex data presented on the development of LH by Maier and Watkins

(2005). However, the approach really is a simplification and one major drawback is that we have not provided any normative arguments for such a generalisation. Generalisation of values according to similarities of states are well-known and form the basis of extensive work on approximation to value functions, e.g. with neural networks (Sutton and Barto, 1998; Bertsekas and Tsitsiklis, 1996) or in hierarchical formulations (Dayan and Hinton, 1993; Dietterich, 2000). However, we are not aware of other instances in which the values are generalised between states that share similar distribution of rewards. While at first sight an intuitive approach, it is easily seen that it has to be applied cautiously.

Consider a point in A in a maze at which the optimal action is to go right, whereas the optimal action at some other point B is to go left, and in both cases the optimal actions lead to apples, whereas any other actions lead to no reward. Our proposal would assign similar action choices to both states just because the optimal action in each of these two states always leads to an apple. The distribution of rewards only carries information about action choice in very particular situations — for example here in LH. However, as in this chapter, it may be argued that extreme reinforcement distributions are more likely to carry information about the appropriateness of *classes* of actions. Finally, there may be questions about the relationship between the slow onset of the escape deficit and generalisation. Willner et al. (1992a) for example suggest the absence of an early escape deficit has to do with an initial phasic excitation due to entering a new environment. Nevertheless, the proposal does reproduce the behaviour and could easily be put to test in behavioural animal experiments.

Also, an arbitrary function $\lambda(t)$ was used to arbitrate between information gained in the present environment e and information gained in previous environments $i \neq e$. It was chosen as an monotonically increasing function of time to reflect the fact that over time, more is known about e, and less will have to be inferred about it from other environments. A more thorough approach would weight the contributions from e and $i \neq e$ by an explicit measure of their certainty. It may be that methods from Bayesian Q-learning (Dearden et al., 1998) might apply, but this has not been explored as yet. One point of note is that both the increasing function $\lambda(t)$ and a Bayesian approach would predict that after prolonged exposure to the shuttle box, animals would start escaping again, even after IS. Although this has not been tested specifically, it does not appear to be the case (Steven F. Maier, personal communication).

3.5.4 Valence generalisation

The simple models of this chapter fail to account for one important aspect of the data: the generalisation across reinforcer valence. It is clear that stress affects the behavioural responses to rewards: both IS and CMS increase ICSS thresholds specifically in the mesolimbic system (Zacharko et al., 1983; McCutcheon et al., 1991; Moreau et al., 1992); impair the acquisition of appetitive tasks (Ghiglieri et al., 1997; Mangiavacchi et al., 2001); there is direct evidence for down-regulation of striatal dopamine D_2 receptor density after CMS (Willner et al., 1992b) (though see also Willner (1991); Di Chiara et al. (1999); Bekris et al. (2005) and Nanni et al. (2003) who argue for alterations to D_1 receptors) and, intriguingly *in*creases in DA release (Stamford et al., 1991). Thus, not only does stress affect behaviour in tasks in which there is no conflict, but it also appears to directly influence the neurobiology of reward systems. However,

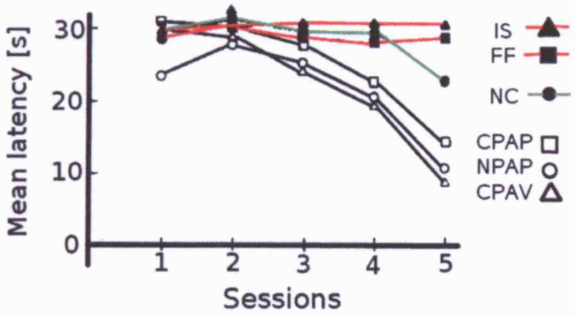


FIGURE 3.13: Escape deficit from free food. The plot shows escape latencies in a shuttle-box task with electric grid floor in six groups of rats that have been exposed to different pre-treatments. Group CPAV were trained to pull a chain to escape aversive stimulation; group CPAP had to pull a chain for food; group NPAP had to nose-poke for food; NC were the naive, home-cage controls. IS were given inescapable shock, yoked to CPAV. FF rats finally were exposed to uncontrollable food, in that they were delivered a pellet whenever their "master" CPAP rat earned one by pulling the chain. There is an escape deficit for animals exposed to uncontrollable contingencies, and this generalised across reinforcer value. Arguably, the master animals show effects of mastery here, as they acquire the escape response more rapidly than the naive controls. Figure adapted from Goodkin (1976).

most crucially, the effects exists in the opposite direction: exposure to uncontrollable appetitive events, such as yoked food delivery, also leads to escape deficits (see figure 3.13; Goodkin 1976; Overmier et al. 1980). To explain such effects in the same framework as used in this chapter, an action that blunts symmetrically would have to be postulated, i.e. which decreases reward sensitivity in parallel to inducing analgesia. To our knowledge, there is no evidence that for example morphine inhibits appetitive behaviours. Thus, results like those in figure 3.13 are one of the major motivations for exploring goal-directed models of learned helplessness in the next chapter.

3.5.5 PREDICTABILITY

We have here addressed some of the effects of controllability. One important avenue for further research is the effect of predictability. In terms of blunting, it is known that analgesia interacts in complex manners with predictability. Lysle and Fowler (1988) find that presentation of an aversive US before testing leads to hypolagesia and negatively conditioned CS produce hyperalgesia. However, given the choice, animals still prefer predicted over unpredicted shock Badia et al. (1979, 1983); Mineka and Hendersen (1985) and produce different stress responses (Dess et al., 1983). Prima facie, predictability is not explicitly part of the cached values. However, as in the present discussion of control, it may be that some of the behavioural effects ascribed to predictability can be generated by a system that has not explicit knowledge of it.

IV

CONTROL

ABSTRACT

At its core, learned helplessness states that subjects infer the degree of their experimental control in an environment and then generalise this knowledge to other environments. It is a controversial concept that has attracted extensive animal and human experimentation. Here we proceed to formalise this notion in a fully Bayesian reinforcement learning (RL) framework. We present progressively more powerful and relevant notions of control: control as a prior on the degree of branching of a decision tree; and as a prior on the extent to which outcomes, and then reinforcements, in an environment are reliably achievable. We present evidence that control has profound consequences for exploration behaviour and motivation; that it is the relevant statistic in a wider RL setting; and that it can account for several important aspects of animal models of depression. Finally, it provides a framework for the formalisation of inter-individual differences in susceptibility to LH and for attributional issues and may thereby yield insights into resilience.

4.1 INTRODUCTION

The previous chapters have detailed how the notion of control has been crucial to theories of depression. Animals display behavioural appreciation of the extent to which reinforcers are under their control. In the animal literature, control is defined as the degree to which a desired outcome, or a reinforcement, can be evoked by taking an action. Animals that are exposed to

environments without such control develop subsequent behavioural traits that model (Willner, 1986; Willner and Mitchell, 2003; Frazer and Morilak, 2005) aspects of depression in humans, in that they respond less to and learn less rapidly about both rewards and punishments (Overmier et al., 1980; Willner, 1997; Maier and Watkins, 2005). Similarly, making it clear to humans that they are not in control of some relevant aspects of their environment increases some measures of depression (Miller and Seligman, 1975; Blaney, 1977; Miller, 1979), but maybe more importantly, expression of such sentiments increases the risk of developing depression (Alloy et al., 1999).

While the human literature does focus on (conscious aspects of) goal-directed behaviour (people are asked about their judged ability to achieve goals, or their beliefs specifically about this is manipulated), the paradigms employed in the animal literature have often been close to habitual ones. This is particularly true for pure LH, which arose from experiments probing theories of habitual learning (Overmier and Seligman, 1967; Seligman, 1975; Peterson et al., 1993). Indeed, in chapter 3, we found, somewhat to our surprise, that extensive aspects of the animal phenomenon can be explained by a combination of blunting and cached reinforcement learning, without any recourse to computational notions of tree search or goal-directedness. This analysis has identified some key aspects that remain beyond the grasp of a habitual, cached learning system. Mainly, this is the finding that LH effects generalise across reinforcer valence (figure 3.13) — rats exposed to IS come to show pure appetitive learning deficits, and exposure to uncontrollable positive reinforcements equally induces an escape deficit (Overmier et al., 1980; Goodkin, 1976).

In this chapter, we formalise this notion of control, or rather, attempt to find a formal definition of control that can account for the major features of the data on learned helplessness, particularly those features unaccounted for by habitual learning. This does not mean that we judge habitual learning not to take part in learned helplessness — or, for that matter, depression — but rather that certain aspects of it may be better described of in a goal-directed framework. Generalisation is at the heart of all experimental tests of control. Thus, it is important that the formalisation should allow for control to be inferred in one environment, and then applied to a second environment (Seligman and Maier, 1967; Maier and Watkins, 2005). Furthermore, we are interested in keeping a strong link to normative arguments, and thus seek a notion of control that makes contact with the issue of generalisation in a broader reinforcement learning setting. For if control is a powerful enough statistic to underlie aspects of psychopathology, then one may expect a role for it in normative setting. For both of these reasons, we will work in a Bayesian framework and propose a description of Markov decision problems parametrised such as to emphasise the effects of various notions of control.

As is natural with respect to the literature (Maier and Seligman, 1976; Abramson et al., 1979; Alloy and Abramson, 1982), the formalisation of control we propose rests within the framework of model-based tree search (Sutton and Barto 1998; Bertsekas and Tsitsiklis 1996) as applied to goal-directed action choice (Daw et al., 2005). There, action sequences are evaluated by explicitly constructing, for each sequence of actions, a probability distribution over the (known) outcomes with the help of an explicit model $\mathcal{M}$ of action-outcome associations (see figure 1.1, and more generally figure A.2). Vanilla tree search is only applicable to small problems (Sutton and Barto, 1998; Bertsekas and Tsitsiklis, 1996) because the number of action sequences needing

to be evaluated grows exponentially as $D^{|A||O|}$, where D is the length of the action sequence, $|A|$ the number of actions available and $|O|$ the number of outcomes for each action. The exponent is known as the branching factor of a RL problem. For large problems, approximate solutions have to be found. Here, we are mainly concerned with the contribution of $|O|$ in scenarios of repeated choice amongst a fixed set of actions.

The formalisation of control proposed here is relevant in situations in which the model $\mathcal{M}$ itself is unknown, but in which observations $\mathbf{N}$ are combined with a prior belief on models $p(\mathcal{M})$ to furnish a distribution over models $p(\mathcal{M}|\mathbf{N})$ that can be used for model-based tree search (Dearden et al., 1998, 1999; Engel, 2005). We argue that control is related to the branching factor of $p(\mathcal{M})$, although in a number of subtle ways.

In the following, we develop our reward-sensitive notion of control in an incremental manner, starting from a simple prior on the entropy of the outcome distribution. The emphasis is on developing these notions of control and their consequences in a RL setting, rather than detailed comparisons with experimental data. We illustrate the relevance of the various notions to different aspects of depression, and to RL in general. Nevertheless, we do return to animal models of depression and illustrate to what extent the formalisations of control can qualitatively account for the data. Finally, we discuss some aspects of the attributional reformulations (Abramson et al., 1978, 1989) of learned helplessness.

4.2 NOTIONS OF CONTROL

Let us first give an overview over the major notions of control we will consider. Throughout, mathematical details are relegated to appendix B.

We consider environments in which a number $|A|$ of actions a is available. In general the actions may be elemental actions (e.g. muscle activations) or more complex actions (e.g. walking, grasping), but here we have in mind even more complex "mega-actions", such as playing a game of tennis, embarking on a PhD programme or getting married. Each action, at all the levels, leads to outcomes. For elemental actions these are basic movements, and for the mega-actions these are more global outcomes. For tennis, the game might be won, lost, result in fun, an ankle may be sprained or the racket broken. Nevertheless, for exposition purposes, we will use a vending machine with $|A|$ different buttons as an example.

The first, most basic notion of control (which was partially formulated by Maier and Seligman (1976) and that underlies the work on "depressive realism" Abramson et al. 1979; Alloy and Abramson 1982; Alloy and Tabachnik 1984; Msetfi et al. 2005) is related to the breadth of different outcomes for each action, or the entropy of the outcome distribution (figure 4.1A). If p_o are the probabilities of the various outcomes, the entropy of the outcome distribution is

$$\mathcal{H} = -\sum_o p_o \log(p_o). \tag{4.1}$$

The entropy of a distribution is the most standard measure of its spread. It measures the "expected surprise", i.e. on average how surprised we are by observations drawn from the distribution (Cover and Thomas, 1991; MacKay, 2003). If $\mathcal{H} = 0$, we are entirely unsurprised, or, put

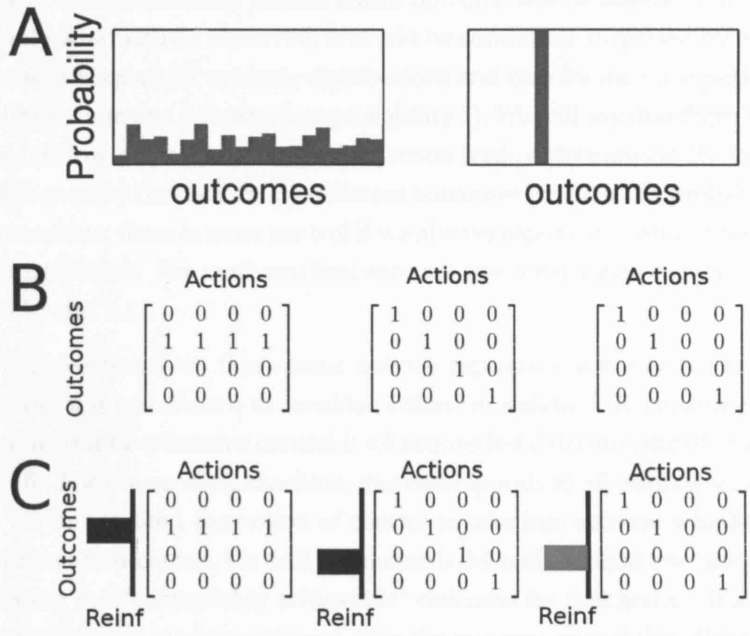


FIGURE 4.1: Notions of control. **A**: Outcome entropy. There is more control if an action only has a few likely outcomes (left) than if many outcomes are likely (right). **B**: Fraction of controllably achievable outcomes. If there is more than one action, the relationship between the outcomes of the different actions is important. In the matrices, each column is related to the outcome distribution of one actions. In this case there are four actions (four columns). Each column has a unity entry on the outcome for which there is a marked peak in that action's outcome distribution. If there is no unity entry, the outcome distribution is flat. Consider the leftmost matrix. All actions only lead to one outcome, but they all lead to the same outcome, like a vending machine which only has one type of chocolate bar to offer. Whatever action is chosen, the same outcome results. On the other hand, in the middle matrix, each action leads to another outcome. Different buttons on the vending machine do yield the different outcomes advertised. There is more control if a different action can be chosen to achieve each outcome in the environment. The rightmost matrix shows a case in between, where the vending machine yields three out of the four outcomes advertised. These three matrices are examples of **M** matrices. **C**: Fraction of controllably achievable reinforcement. There is most control if rewards are under behavioural control. The bars represent the (positive) reinforcement associated with each outcome. In the left case, all reinforcement is associated with the only possible outcome for all actions. All vending machine buttons yield the one chocolate bar we desire. In the middle matrix, there is one button which yields to the desired bar, the others yield to other outcomes. We say that in these cases there is full control. However, if the reinforcement is as indicated by the red bar in the right matrix, then all but the reward-carrying outcome can be achieved. We can get all kinds of bars but the one we want. In this case there is no controllably achievable reward.

differently, we can always correctly predict which outcome will be observed. If $\mathcal{H} > 0$, we cannot predict exactly what will be observed, and will be somewhat surprised by every outcome. The entropy is thus maximal for uniform distributions and zero for the most peaky distribution (in discrete settings when one outcome has probability 1). We will say that there is more control when an action has low outcome entropy (if an action leads deterministically to one outcome) than if it has high entropy (leads to many different outcomes with similar probability). In terms of the vending machine, there is more control if we always receive the same chocolate bar when we press the same button. For mathematical ease we use constrained outcome distributions, see equations B.13 and B.14.

While we will see that even these basic notions reproduce some of the fundamental results, it is clear that it is insufficient to consider actions in isolation. According to the previous definition, there would be extensive control if all actions led deterministically to the same outcome (figure 4.1B). For the vending machine, this corresponds to all buttons yielding the same chocolate bar. We thus extend the notion of control to take into account whether different actions achieve different outcomes. We will use a matrix $\mathbf{M}$ with at most one unity entry in each column designating the "controllably achievable" outcome for that action. If a column of the matrix $\mathbf{M}$ has a unity entry at some outcome, then the outcome probability distribution for that action is peaked at that outcome. The total number of ones in the matrix, $|M|$, then designates the number of outcomes that are controllably achievable within an environment. If there are L possible actions, $|M|/L$ (with $|M| < L$) is the "fraction of controllably achievable outcomes". When this fraction is one or greater, it is possible to choose an action to achieve each outcome in the environment.

Finally, consider a situation in which there is one predominant need and there are actions available which deterministically lead to every kind of outcome, but no action reliably leads to the fulfilment of the predominant need. For example, we might want a particular chocolate bar from the vending machine, and the buttons only yield all kinds of bars and sweets but the one we desire. Another good example for this is the yoked rat in the learned helplessness model, which may be able to do all kinds of things, but which cannot switch off the shock (figure 4.1C). We will thus say that control is larger the larger the fraction χ of reward which can be earned from outcomes that are controlled by some action. We will call χ the controlled reinforcement.

4.3 Results

This section discusses the three notions of control in sequence. It explores the consequences of each and then illustrates the applicability to two animal models of depression: the classical learned helplessness model (Maier and Seligman, 1976), and the chronic mild stress model (Willner et al., 1987; Willner, 1997).

4.3.1 Outcome entropy

The simplest notion of control — that of outcome entropy — has important repercussions on exploration behaviour, the average expected reward and the incentive contrast between actions.

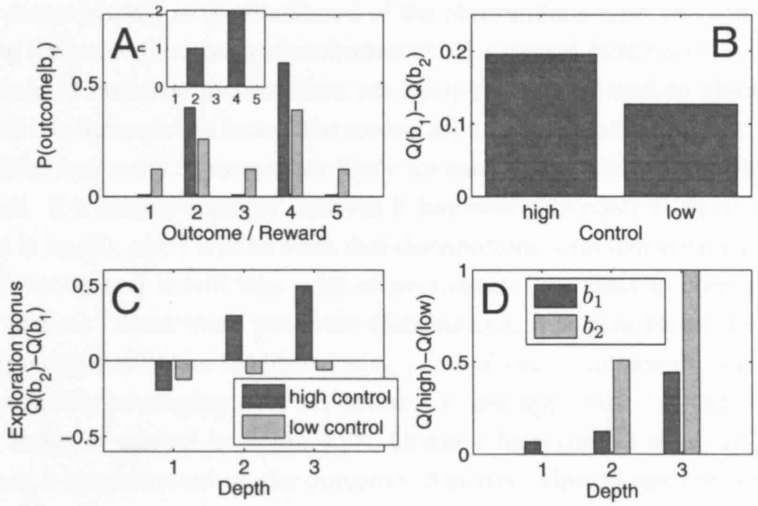


FIGURE 4.2: Effect of outcome set size prior on $\mathcal{Q}$ values and exploration. **A**: After observations $\mathbf{n}$ in the inset for pressing button 1 three times, the dark bars show the predictive distribution for button 1 with, and the light bars the predictive distribution without control. With control, it is unlikely that outcomes are observed that have not yet been seen, and thus the predictive distribution will always have lower entropy than without control, and be peaked on the outcomes already observed. **B**: Expected immediate reward under high- and low-control priors. **C**: Exploration bonus: difference between the $\mathcal{Q}$ values of the unknown (b_2) and known (b_1) button for depths $D = 1, 2, 3$ with (dark bars) and without (light bars) control. **D**: Difference between the $\mathcal{Q}$ values of each action under high- and low-control priors.

It can be interpreted as the degree of branching of a decision tree. In terms of our vending machine analogy, we imagine having the option to try a number D of buttons. The outcome entropy will determine how many different buttons we will try out, how much we expect to get overall, and how much we will prefer one button over others. Mathematical details are presented in appendix B.1.

Consider choosing between two unlabelled buttons on the vending machine, b_1 and b_2, each of which has $L = 5$ possible outcomes, with outcome o yielding reward $R_o = o$. Assume button b_1 has been tried three times already, with the outcomes displayed in the inset of figure 4.2A, but that nothing else about it is known. Button b_2 has never been taken and nothing is known about its outcomes. The most advantageous button to press is the one with the highest expected reward. The expected reward for button b_1 is simply $\sum_o c_o^{b_1} R_o$, where $c_o^{b_1}$ is the probability of observing outcome o after pressing button b_1. The true value for the two actions cannot be calculated because the true outcome probabilities c_o is unknown for both, but given observations (in fact just a count of outcome frequencies $\mathbf{n}$), a posterior distribution over the outcome probabilities of some action a can be derived by combining the observations with a prior according to Bayes' rule:

$$p(\mathbf{c}^a|\mathbf{n}) \propto p(\mathbf{n}|\mathbf{c}^a)p(\mathbf{c}^a) \tag{4.2}$$

Here, the first factor $p(\mathbf{n}|\mathbf{c}^a)$ is the likelihood of the observations $\mathbf{n}$ on one action given some true underlying (unknown) outcome distribution $\mathbf{c}^a$. The second factor $p(\mathbf{c}^a)$ is the prior belief about what kinds of outcome distributions are likely (o buttons tend to yield one or many outcomes?), and it is through this factor that control is implemented throughout. In this section, this prior specifies *how many* outcomes are likely for each action, i.e. the prior is on the size of the outcome set. If a subject strongly believes it has extensive control (the outcome set size for each action is small), $p(\mathbf{c}^a)$ will be such that distributions with low entropy are inherently much more probable, and it will take a lot of persuasion from data to convince the subject that it has no control. From these posterior distributions, it is possible to derive predictive distributions (predictions of the likelihood $p(n_{D+1}|\mathbf{n})$ of each outcome on the next choice of the action, derived by averaging over all possible $\mathbf{c}$, see appendix B.1 and B.1), shown for high and low levels of control in figure 4.2A. Under a high-control prior, all the predictive probability mass is concentrated on the outcomes that have already been observed, while for a low-control prior the predictive distribution is broader: if we believe that a button produces just one outcome, then after observing a particular outcome from that button, we will predict that that outcome will always be observed. Thus, control here only affects *which* outcomes are predicted, not what their associated reward might be.

The various consequences of the predictions are displayed in the rest of the figure. Action b_2 has never been tried, so its predictive distribution is flat and the expected outcome of it is 3. Because outcome 4 was observed twice, and outcome 2 only once, the expected reward of action b_1 under both high and low-control priors exceeds that of action b_2, more so in the high- than in the low-control situation (figure 4.2B). But this is the case only if a single action choice remains (for a decision tree of depth $D = 1$). If two or more actions remain to be taken, it becomes worth trying out the unknown button b_2 to ascertain whether it might not really be better than b_1. This phenomenon, whereby it becomes advantageous to try out unexplored actions, rather than exploitatively choosing the best known action, is called an exploration bonus. The exploration bonus is a function of control. To see this, imagine that button b_2 were chosen and yielded outcome and reward 5, i.e. some absolutely delicious chocolate bar. Under the high-control prior, the predictive distribution will now be strongly peaked on outcome 5, and we will choose button 5 again for our second action choice. Under the low-control prior on the other hand, this individual outcome affects the predictive distribution very little and the value of action b_1 remains superior. The nice outcome is assigned to pure chance, and it is not worth changing one's course of action. Thus, under high-control priors, not only are actions that lead to good outcomes aggressively exploited (and actions with negative outcomes equally avoided), but the option of exploitation makes exploration worth the while. The opposite is true under low-control priors, where outcomes bias action choice only weakly and where there is no exploration because any good outcomes are not assumed exploitable. Figure 4.2C shows the $\mathcal{Q}(b_2) - \mathcal{Q}(b_1)$ for one, two and three remaining action choices. Not only is the sign of the difference between the actions different (indicating presence *vs* absence of an exploration bonus), but the absolute size of the difference is also bigger. Thus, under high-control priors, there is more contrast between actions, and this increases when more actions remain to be chosen. Figure 4.2D shows this more explicitly for this toy example. Furthermore, because rewards are assumed exploitable and punishments avoidable, the expected average reward under high-control priors is always greater (at worst equal to) that under low-control priors.

4.3.2 FRACTION OF CONTROLLABLE OUTCOMES

So far, we have only looked at actions in isolation. A more global notion of control should take into account to what extent different outcomes in an environment can be controlled individually. The second definition of control is accordingly the fraction of controllably achievable outcomes $|M|/L$, where $|M|$ is the total number of outcomes for which an action exists that produces it with high probability. In figure 4.1B, $|M|$ is 1, 4 and 3 for left, middle and right panels respectively, and $|M|/L$ is $1/4$, 1 and $3/4$. How does this enhanced notion of control affect action choice? First note that it is an extended version of the first notion, in that there is still no control if the outcome entropies for all actions are large (the precise relationship between outcome entropy and $|M|$ is detailed in appendix B.2.1).

For the vending machine, we thus define control as the fraction of advertised bars that can be obtained reliably by pressing a particular button.

To illustrate the differences with the previous setting that neglected relations between actions, imagine an vending machine with $|A| = 5$ buttons, $L = 5$ possible outcomes and in which we are allowed to press $D = 4$ buttons. Pressing button o preferentially leads to outcome o. Only one button (button 1) has ever been taken before (4 times) and it has always yielded outcome 1 with reward 0. Let the rewards for the outcomes be $R = [0\ 0.2\ 0.23\ 0.27\ 0.3]$ which has the property that the expected value of unexplored actions ($1/L \sum_o R_o = 0.196$) is just smaller than the reward associated with the second action. The four other buttons a_i result in outcome $o = i$ with highest probability. Button 5 is therefore, unbeknownst to us, the best action. For illustration, let us force exploration to proceed in an ordered manner, from action 1 to 5, i.e. if we decide to try a new button we have to try the next one in the sequence — we can't just jump ahead and try button 5 (there is also no reason why we should want to, given that we know nothing about either of the buttons 2-5). Then, the exploration depth — the action at which exploration ceases — is a measure of the degree of exploration "drive". Figure 4.3A shows the consequences of different priors on the exploration depth. Priors are hard and only allow predictions consistent with M matrices of a particular $|M|$.

- $|M| = 0$: We believe that no button will reliably lead to any outcome. Even after observing the first outcome 4 times, the predictive distribution is flat for all actions, *including* button 1. Thus, the button 1 looks as good as all other buttons about which no information has been gathered. Figure 4.3A shows that all four draws for a prior that enforces $|M| = 0$ result in the choice of button 1.

- $|M| = 1$: We believe that one button will reliably lead to one of the outcomes. The ML estimate of M has its only nonzero entry on button 1 and outcome 1, all other buttons are assumed to generate any of the L outcomes randomly. Due to our choice of R, button 2 is advantageous over button 1. Thereafter, button 2 will be chosen, as its outcomes (outcome 2 with $R_2 = 0.2$) are marginally larger than those from the unknown actions. Figure 4.3A shows that all four actions for a prior that enforces $|M| = 1$ result in the choice of button 2.

- $|M| = 2$: We believe that two buttons will reliably lead to two different outcomes (one outcome each). Again, button 2 looks better than button 1 for the first action choice.

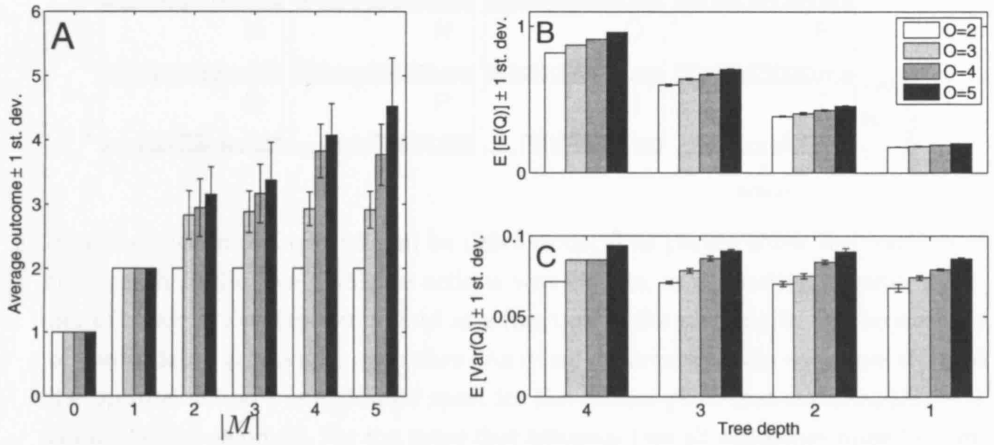


FIGURE 4.3: Effect of prior belief about fraction of controllable outcomes on exploration and expected rewards. **A**: Exploration. $D = 4$ remaining actions, with $A = L = 5$ actions/outcomes. The bars show which action is taken on the first (white, $D = 4$), second (light grey ($D = 3$)), third (dark grey ($D = 2$)) and fourth (black ($D = 1$)) trials on average over many trials. The bar groups for priors putting exclusive mass on $|M| = \{0, 1 \cdots 5\}$ controllable outcomes. As more outcomes are assumed to be controllably achievable, exploration proceeds further. Action 5 was reached 40% of the time when the prior assumes that all actions are controllably achievable ($|M| = 5$), and only 5% of the time when $|M| = 4$. However, for $|M| = 5$ the variance of the third and fourth trials are large as well because a spurious high reward on one of the other actions (which here occurred in 2/10 trials) leads to exploitation of that action. **B**: Q values for priors peaked on $|M| = \{2 \cdots 5\}$. Bars show the mean of the average Q values across all states, over all trials. The error bars indicate the standard deviation over trials. In all cases, a prior that assumes larger fraction of controllably achievable outcomes on average leads to higher expected rewards. **C**: Variance of the Q values across states. Bars indicate mean variance, error bars indicate standard deviation of the Q-value variance over trials. A high control prior leads to larger differences between the value of actions — a larger incentive contrast between actions.

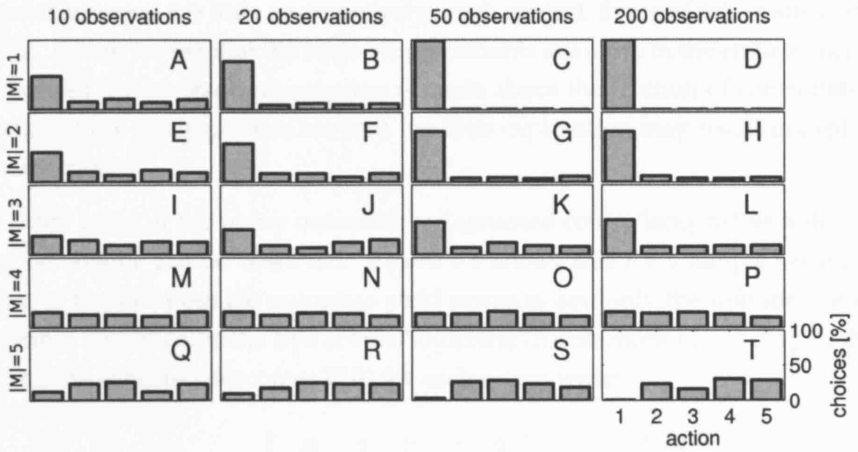


FIGURE 4.4: Too much control can be deleterious. The panels show the fraction of times each of the five available actions was chosen, as a function of prior number of observations (columns) and as a function of the prior belief on the number of controllably achievable outcomes. As more observations are used, we see that the optimal action 1 is exploited most for the correct prior that assumes $|M| = 1$ controllable outcomes. For the prior that assumes that all outcomes must be controllable ($|M| = 5$), we contrarily see that action 1 is avoided.

Thereafter, however, there is a chance that the second nonzero entry is assigned to button 3 / outcome 3. The predictive distribution for button 3 will not be flat, and thus the expected outcome for that button will be greater than the expected reward for button 2. However, exploration will mostly stop at button 3, as shown by the set of columns in figure 4.3A for $|M| = 2$.

- As the matrix $\mathbf{M}$ is constrained to contain more nonzero entries in different columns, i.e. as we believe more and more of the outcomes are achieveable through some button, exploration proceeds until all buttons have been explored.

These exploration effects are due to a graded analogue of the effects shown in figure B.2. Figure 4.3B and C also show that, similarly to the previous setting, the average Q values increase as the priors put more mass on larger $|M|$, and that larger $|M|$ mean actions differ more in their expected rewards.

4.3.3 GENERALISATION

Assume two environments share levels of control, and the level of control is inferred in one of them. Is it advantageous to generalise this inferred statistic to the other environment about which otherwise nothing is known? That is, does accurate knowledge about the level of control in an environment help action choice? Is it the case, that knowing about the number of buttons which will reliably produce one outcome will help us make better choices? It is only useful to generalise a parameter if knowledge about that parameter indeed confers advantages.

In support of this hypothesis we find that performance in a new environment degrades both

when the prior assumes too little or excessively much control. Some of this is already apparent in figure 4.3A, where we see that the large reinforcements available in the environment are only effectively reaped if the correct assumption is made about the fraction of controllably achievable outcomes: if the assumption is too low, too little exploration may result in exploitation of a suboptimal action.

On the other hand, if too many outcomes are assumed controllably achievable, a similarly suboptimal behaviour can be observed. Figure 4.4 shows this for a simple setting in which only two out of the five possible outcomes yield rewards, and only the inferior one is controllably achievable. Specifically, the five actions' outcome distributions $\mathbf{C} = \{\mathbf{c}^a\}_{a=1}^5$, the reward vectors $\mathbf{R}$ and the expected outcomes $\mathbb{E}[R]$ for each action were:

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} .8 & .2 & .2 & .2 & .2 \\ .05 & .2 & .2 & .2 & .2 \\ .05 & .2 & .2 & .2 & .2 \\ .05 & .2 & .2 & .2 & .2 \\ .05 & .2 & .2 & .2 & .2 \end{bmatrix}; \quad \mathbf{R} = \begin{bmatrix} .3 \\ 0 \\ 0 \\ 0 \\ .7 \end{bmatrix} \\ \mathbb{E}[\mathbf{R}] &= \begin{bmatrix} .275 & .2 & .2 & .2 & .2 \end{bmatrix} \end{aligned} \tag{4.3}$$

which means that the matrix $\mathbf{M}$ corresponding to the matrix $\mathbf{C}$ has only one unity entry in the top left corner, making the true $|M| = 1$. From the reward vector $\mathbf{R}$ we see that only outcomes 1 and 5 carried rewards. As action 1 controllably achieves outcome 1 80% of the time it is the optimal action, despite leading to the suboptimal reward. In terms of the vending machine, this means that button 1 yields a bar which we desire with measure 0.3 80% of the time, and 5% of the time it yields the bar we desire most (0.7). All other buttons yield chocolate bars at random, with three of them that give us no reward at all. Despite not reliably giving us the best bar, button 1 on average satisfies our desire most.

For a varying number $N = \{10, 20, 50, 200\}$ of trials, observations were randomly generated from random action choices , i.e. for each trial a random action was chosen, and for that action a random outcome generated. The posterior and predictive distributions given this data and the various priors were then evaluated, and two more actions chosen (because two is the smallest number required to show an explicit exploration bonus). The prior distribution allowed $|M|$ controllable outcomes, i.e. it allowed matrices $\mathbf{C}$ that were consistent with $|M| = 1$ (figure 4.4A-D), $|M| = 2$ (figure 4.4E-H) etc. Figure 4.4A-D shows that a correct assumption of only one controllably achievable outcome leads to the exploitation of action 1. As more outcomes are assumed achievable, there is more persistent exploration, and this swaps over when all outcomes are assumed achievable and action 1 ends up being *avoided* despite being the optimal action. The pattern becomes clearer when more prior observations are used to infer the predictive probabilities (rightmost column, figure 4.4D and 4.4T), but is already apparent after few observations (on average two per action, leftmost column). As long as the maximal reward is not exploitable, an assumption that more outcomes are controllably achievable than is actually the case will lead to persistent exploration and prevent adequate exploitation. Of course, the cost incurred by this policy will depend on the difference between the amount earned for explorative actions compared to exploitation of the sub-maximal reward.

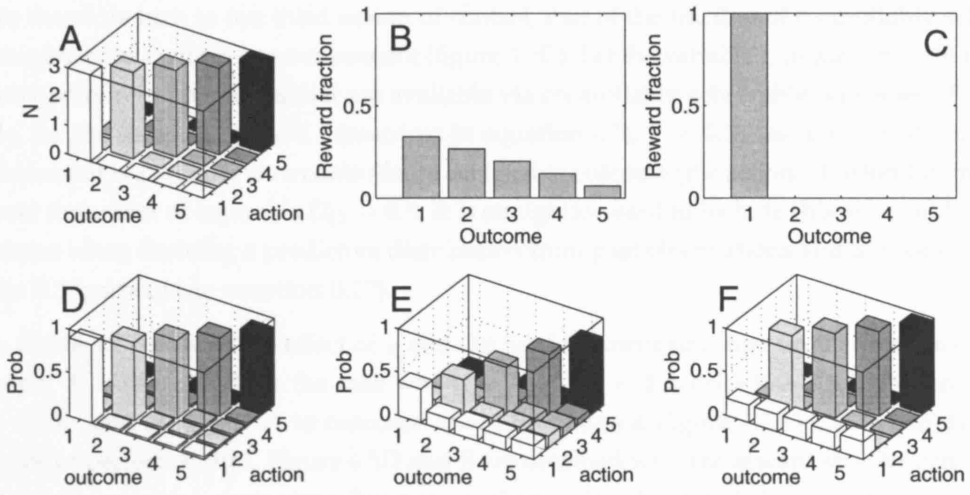


FIGURE 4.5: Reinforcement-sensitive control. **A**: For each action a, outcome a was observed 3 times. **B**: Reward fractions for each outcome as used in figures D and E. Here, all outcomes, and thus all actions, carry sizeable reinforcements. **C**: Reward fractions as used in panel F. One outcome carries all the reward. **D-F**: Inferred action-outcome matrices. Because the tree is constructed from repeated choices, these are also inferred transition matrices. **D**: With the assumption that a large fraction of the rewards in panel B are controllably achievable ($\chi = 1$), predictive distributions $p(n_{D+1}|\mathbf{N}, \chi)$ of low-entropy are recovered for all actions. **E**: However, when $\chi = 0$, the predictive distributions all have a high entropy, and more so the higher the reward of the outcome associated with the action. The rewards here are still those from panel B. **F** The more extreme reward distribution of panel C, combined with a $\chi = 0$ results in a predictive distribution that has low entropy for the actions that do not lead to rewards, but a high entropy for the one action that leads to the only reward available in this environment. Throughout, $\sigma = 0.05$. Smaller σ accentuate the effects further. See equation B.26 for definition of σ.

Thus the controllably achievable fraction of outcomes is an informative parameter, that is it is of advantage to use the control parameters that describe the environment best, and it is consequently advantageous to generalise this parameter whenever environments share it.

4.3.4 REINFORCEMENT-SENSITIVE CONTROL

However, the number of outcomes seems to be an insufficient metric on which to measure control. Not only is this hard to define in more natural environments, but in terms of depression, the link to reinforcements seems incomplete: it is not the crude number of uncontrollable outcomes that matter (depressed people are perfectly able to exert control over most aspects of daily functioning), but it is control over those outcomes associated with most reinforcement that is relevant. We do not care about how many random unwanted chocolate bars we can reliably obtain from the vending machine — we only care about the one we desire.

We therefore turn to our third notion of control, that of the fraction of controllably achievable *reinforcements* within an environment (figure 4.1C). Let the variable χ (equation B.25) index the fraction of reinforcements that are available via controllably achievable outcomes. For example, for the setup in figure 4.4 (matrices in equation 4.3), $\chi = 0.24$, as only 0.3 of the total reinforcement is available via a controllably achievable outcome (the action / button 1 in matrix **M**), and the extent of control is $C_{11} = 0.8$. It is straightforward to include this as an additional constraint when deriving a predictive distribution from past observations and a prior (see appendix B.3, particularly equation B.27).

In figure 4.5 we show the effect of χ and the reinforcement structure on the predictive distribution. It is illustrated for the case where each of $|A| = 5$ actions has already been taken three times, and always lead to outcome $o = a$ for action a (figure 4.5A), i.e. when there is evidence of perfect control. Figure 4.5D and E are obtained with the reward structure in panel B, where all outcomes carry some, but not equal amounts of, reward. In panel D, $\chi = 1$, and thus only matrices **M** that have one unit entry in each column and correspondingly outcome probability vectors $\mathbf{c}^a$ of low entropy are allowed to contribute to the predictions. Overall, thus, a very low-entropy predictive distribution is recovered for all actions as all actions carry rewards. However, when χ is set to zero, the predictive distribution changes: the entropy of the outcome distributions for all actions is increased, as all actions lead to rewards, but this is most pronounced for the actions leading to the largest rewards, here action 1. Figure 4.5F shows a more extreme version of this when action 1 is the only action leading to reinforced outcomes. Now all actions are predicted to lead to outcomes deterministically, apart from the one action which produces rewards.

4.3.5 Animal models of depression

Let us now apply our general findings to the two main animal models of depression, learned helplessness and chronic mild stress. At the outset, it is important to note that all our results so far generalise directly to punishments, although they have been phrased in terms of rewards. To see this, we note that all reinforcement vectors can be modified by writing

$$\tilde{R}_i = R_i - \max_j R_j. \tag{4.4}$$

Our definition of fractional rewards already includes such a transform (see appendix B.3). Maximisation of the rewards is now equivalent to minimisation of the punishments.

4.3.5.1 Learned helplessness

In LH, as described in detail in chapter 1, animals are exposed to electric shocks. In the escapable scenario, one action (usually turning a wheel) allows them to terminate the electric shock. In the inescapable scenario, the wheel is taped and cannot be turned. The animals are then transferred to a different, bipartite box. Again, shocks come on at random times (now always delivered via an electrified grid floor), but the escape action is different: rather than turning a wheel, they have to shuttle from one partition to the other.

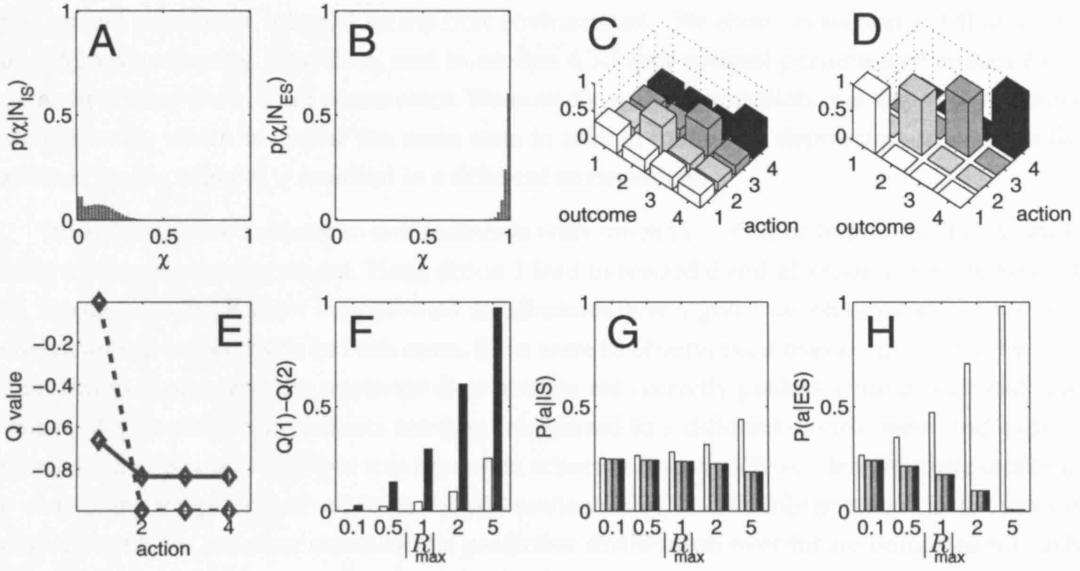


FIGURE 4.6: Learned helplessness after acute severe shock. LH was simulated by first inferring distributions over χ in one environment and then using this as a prior over χ in a second environment. In the first environment, all but action 4 were punished, in the second environment, all but action 1 were punished. **A**: posterior distribution over controllably achievable reinforcement χ $p(\chi|\mathbf{N}_{IS})$ given 80 observations $\mathbf{N}_{IS}$ in a low-control ($\chi = 0.1$) environment in which inescapable shocks (IS) are presented. The distribution is concentrated on low values. **B**: posterior distribution $p(\chi|\mathbf{N}_{ES})$ given 80 observations $\mathbf{N}_{ES}$ in a high-control ($\chi = 0.9$) environment in which escapable shocks (ES) are presented. **C** and **D**: Predictive distributions over outcomes for each action in the test environment. For each action a in the test environment, outcome a was observed 20 times. This is strong evidence for full control. When the low-control prior over χ from panel A is used, this results in high-entropy predictive distributions, but it results in low-entropy predictive distributions if the high-control prior is used. **E**: Q values of the four actions in the test environment. The best action (action 1) has smaller expected reward after exposure to uncontrollable reinforcement (solid line) than after exposure to controllable reinforcement (dashed line). The difference between the actions is attenuated by exposure to uncontrollable rewards. **F**: Increasing the size of the punishment in the test environment has more drastic effects on the advantage of action 1 over the other actions after exposure to controllable than uncontrollable reinforcers. Dark bars show difference between the Q value of action 1 and action 2 after ES, light bars after IS. **G** and **H**: Using the Q values to derive a probabilistic policy. Preference for action 1 (white bar) over other actions (light grey to dark grey bars) increases faster with increasing reinforcer strength after controllable (H) than uncontrollable (G) reinforcement.

We model this by first inferring a control parameter in one environment, and then looking at the acquisition of the correct escape response in a second environment using the setting of the control parameter inferred in the first environment. We show in section B.3 that χ can be inferred accurately from data, and in section 4.3.3 that optimal performance ensues from correctly setting the control parameters. We now show that acquisition of a response in a new environment, which is one of the main tests in animal models of depression, is profoundly affected by the value of χ acquired in a different environment.

We either expose subjects to environments with much ($\chi = 0.9$) or little ($\chi = 0.1$) controllably achievable reinforcement. Here, action 1 lead to reward 0 and all other actions to reward -1. Figure 4.6A and B show the posterior distributions over χ given the observations in the two environments respectively. In both cases, there were 80 observations overall, generated by random action choice, and the posterior distributions are correctly peaked around high and low values of χ respectively. Subjects are then transferred to a different environment and experience a further 80 outcomes, but this time each actions a leads to a fixed, deterministic outcome $o = a$. Using the prior derived from the first environment, and the observations in the second environment, we can now calculate the predictive distribution over future outcomes for each hypothetical value of χ and average over the distributions in figure 4.6A and B. Figure 4.6C shows that when the distribution from figure 4.6A is used, the predictions have high entropy, while figure 4.6D shows that the distribution from figure 4.6B have low entropy. As previously, we can use these predictions to find the Q value of each action. Figure 4.6E shows that action 1 has much higher value after exposure to controllable reinforcements; that the difference between actions is larger; and that the average value is higher (not shown). The second point is explored in more detail in panel F, which shows the difference between actions 1 and 2 as a function of the shock size of actions 2-4. As expected, the impact of an alteration of shock size on the Q values is greater after exposure to escapable than inescapable shock. Figure 4.6G and H finally show the action choice probabilities, again as the shock size is varied. Just as for the difference between the Q values of actions 1 and 2, we see the that differences in choice probabilities grows more rapidly after controllable shocks. After extensive exposure to the controllable test environment, the differences between the groups vanish (not shown), because there is continued learning about χ.

Importantly, this replicates the result by Jackson et al. (1978), whereby an increase in shock size reinstates escape behaviour even in inescapable shocked animals. Imagine shocks of size 5 had been given in the escape task. The escapably shocked animals are at limit. Increases in shock strength will not increase the probability that they choose to escape, but it will increase the probability that the inescapably shocked animals will do so. It also replicates the generalisation finding by Maier and Watkins (2005) (discussed in chapter 3) to a certain degree: Initially, subjects will choose actions randomly, not knowing which outcomes they lead to. Even after being given good evidence that they can escape the shock, they will give little preference to the escape (figure 4.6A and C).

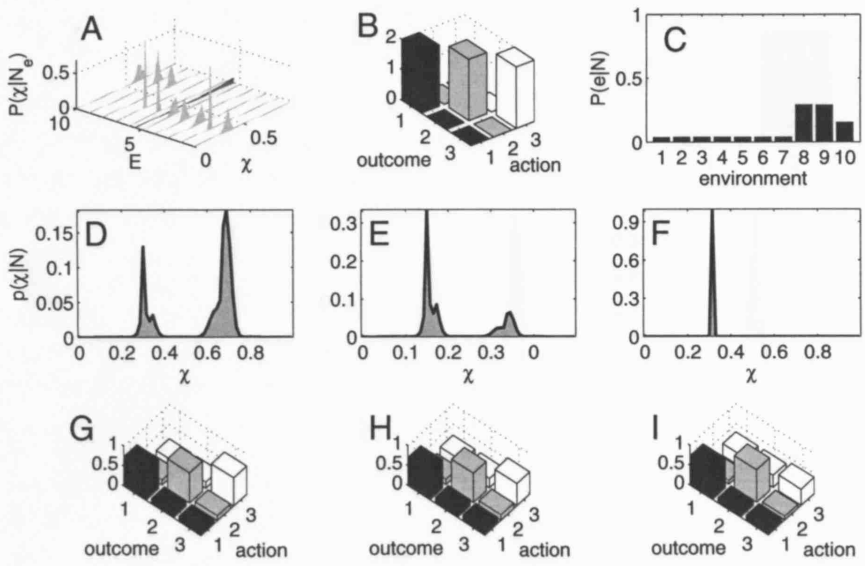


FIGURE 4.7: Chronic variable mild stress with low control χ in 7/10 environments. **A**: light grey: posterior distribution $p(\chi|\mathbf{N}_e)$ for each of the 10 environments. The dark grey histogram shows the posterior $p(\chi|\mathbf{N})$ for the test environment and is nearly flat as few observations have been made. **B**: observations made in the test environment so far. As usual, they are consistent with perfect control. Only outcome 3 is associated with reward in the test environment. **C**: Probability of being in environment e given data $\mathbf{N}$ observed in the test environment. Even a few observations suffice to infer that the test environment has large χ. **D-F**: Priors over χ for the test environment, derived according to three different assumptions about the inter-relationship between χ's of the 10 environments and the test environment. D: mixture weighted by evidence; E: equal mixture; F: product. **G-I** Predictive distributions given observations in panel B and the priors in panels D-F. Only in I is there a sign of helplessness (the predictive distribution for action 3 is flat, while the other two are highly peaked).

4.3.5.2 CHRONIC MILD STRESS

In chronic mild stress, animals are exposed to mild stressors throughout a several-week long schedule, rather than to severe stressors for a small amount of time. This results in escape deficits, decreased primary reward sensitivity and impairments in appetitive learning which are argued to approximate depressive patterns more closely than the behavioural changes after IS. Work in this field has established that the mild stressors have to be varied and uncontrollable — constant or repeated mild stressors do not produce behavioural deficits (Willner, 1997; Cabib and Puglisi-Allegra, 1996). While the construct validity of CMS is very good (section 2.9.1; Willner 1997), its construct is inherently less theoretical than that of LH, and has not generally been thought of in terms of control.

One way of modelling chronic *variable* mild stress is by long exposure to a succession of

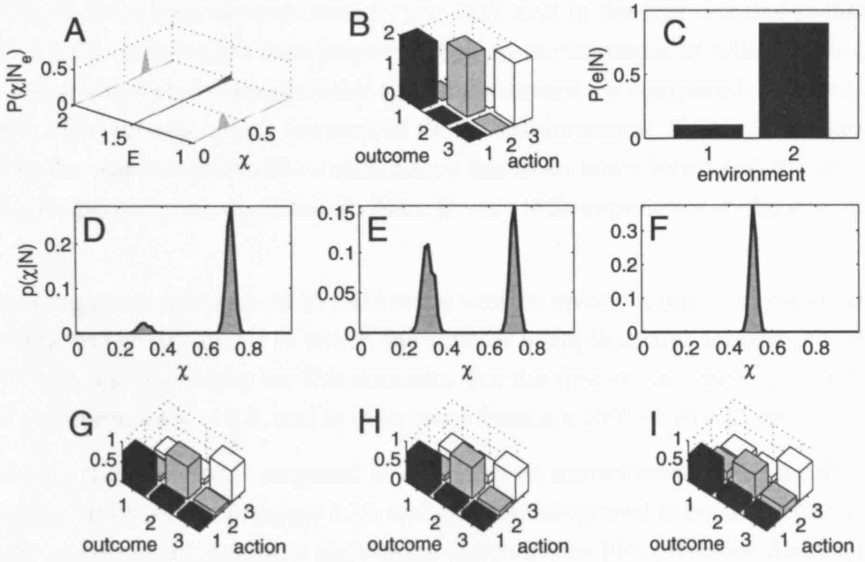


FIGURE 4.8: Chronic repetitive mild stress with low control χ in 1/2 environments. The panels show the same information as in figure 4.7, just for the repetitive stress case. **A**: posterior distribution $p(\chi|\mathbf{N}_e)$ for the two well-known environments home cage and stress cage are in light grey (at ticks 0 and 1). The distribution for the test cage is in dark grey (at tick 1.5). **B**: Observations in test cage so far are consistent with perfect control. Again, all reward is available through outcome 3. **C**: Likelihood of observations in test cage given those in stress (1) and home (2) cage. The test cage is much more similar to the home than the test cage in terms of controllability. **D-F**: Prior distributions over χ derived from different combinations of the posterior distributions $p(\chi|\mathbf{N}^e)$ in panel A. **G-I**: predictive distributions corresponding to the observations in panel B and the respective priors in panels D-F. Weighing the environments by their similarity to the test cage results in a prior with most mass on high χ (**D**) and no signs of helplessness (**G**). Averaging over the posteriors results in a bimodal distribution with equal mass on control and no control (**E**) and no signs of helplessness (**H**). Taking the product of the two distributions results in a medium level of control (**F**), but still with little signs of helplessness (**I**).

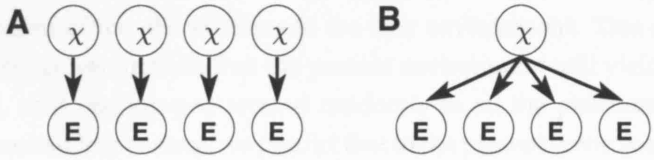


FIGURE 4.9: Generalisation across many environments. **A**: Each environment **E** has a different, private setting of the control variable χ. The overall observation of control will then follow a mixture distribution. **B**: All environments share one single setting of the control variable χ, for example because χ is a descriptor of the single agent present in all environments.

many (10) environments (500 observations per environment). In the majority of the environments (7/10), there is little reward control ($\chi = 0.3$), and in the few remainder there is mild control ($\chi = 0.7$). Subjects are then transferred to an environment in which $\chi = 1$, and the predictions after a few observations in this nice environment are compared. Figure 4.7A shows the posterior distributions over χ for each of the 10 environments. Figure 4.7B shows the observations in the test environment — each action has been taken twice and the outcomes are consistent with perfect control, although there is very little experience in the environment so far.

Most investigations of the *repetitive* mild stress version involve a total of three environments: the home cage, the environment in which the stressor is applied, and some test environment. Figure 4.8A and B show results for this scenario. For the first environment $\chi = 0.7$, while for the second environment $\chi = 0.3$, and in both cases there are 1000 observations.

How should the knowledge acquired in a number of environments (specifically the distributions over levels of control in figure 4.7A and 4.8A) be integrated to predict what will happen in some new environment, i.e. how should the observations (the posterior distributions over χ) be combined to produce a *prior* over χ? We will consider three approaches. This question is inherently very similar to that explored in section 3.4, but we explore it in some more detail here.

1. It is possible to compare the observations in the present environment with those in all previous environments, and weigh the predictions of each previous environment by how likely the observations in the present environment would have been there. This corresponds to the assumption that the present environment is like one of the already observed ones, but there is ignorance about which one (figure 4.9A). This is the same approach as in the generalisation part of section 3.4, and is formally known as a mixture distribution. Figures 4.7C and 4.8C show the weights assigned in this manner to each of the environments seen before the test environment. Note that most weight, even after a few observations (2 for each action), is given to those environments that had high χ. As a result, the prior over χ that will be used for predictions in the test environment (see previous section) is bimodal (figure 4.7D), but the mode corresponding to large χ is much larger than that corresponding to low χ. Therefore, the predictions in figure 4.7G and 4.8G show little if any sign of the exposure to many environments with low control.

2. The prior could be written as an average of the distributions over χ, and be kept that way independent of the observations in the new environment. This corresponds to the somewhat strange assumption that the present environment will yield observations that look, overall, as if one jumped around randomly in all the previous environments. If 7/10 outcomes had low control, we predict that in the present environment 7/10 times the outcomes will be consistent with low control. If many distributions were characterised by a particular χ then there will be a strong prediction that a new environment also has that same value of χ. Figures 4.7E and 4.7H show the resulting distribution over χ predictions for the variable stressor regime. Figures 4.8E and 4.8H do the same for the repetitive regime. Thus, a signature of helplessness starts to emerge in the variable scenario, but not in the repetitive one.

3. Finally, the prior could be written as a product of all the previous distributions over χ. This would correspond to the assumption that really there is only one χ which holds for all environments (figure 4.9B). Taking products of the light grey distributions in figure 4.7A and 4.8A results in the peaked distributions in figure 4.7F and 4.8F respectively. The predictive distributions over outcomes in figure 4.7I now do show the signs of exposure to uncontrollable stress, but the product of the two distributions in the repetitive case results in a distribution midway between the two (figure 4.8I), and thus in relatively little change to the predictions — no helplessness.

Thus, particular aspects of generalisation of control may account for the effects of chronic mild stress, and the distinction between the variable and the repetitive implementations. Helplessness only ensues when the observations from all the environments are combined *as if* there were only one environment, in which case observations from the few environments with control are simply overpowered by observations from the environments without control. We will discuss the relation of this assumption to the internal / external attributional hypotheses (Abramson et al., 1989) in the discussion (section 4.4.4).

4.4 DISCUSSION

4.4.1 NORMATIVE MODEL OF CONTROL

The objective of the thesis is to approach psychiatry in a normative affective decision making framework. As in chapter 3, the account given here is one that relies on optimality. We essentially just started from the Bayesian principle that subjects ought to use all information as efficiently as possible. We used standard probabilistic techniques to draw the most informative (and thus arguably the best (Jaynes, 2003)) conclusions and base action choice upon them, and were led to effects observed in the animal literature. Thus, what are termed models of disease are arguably optimal reactions to events in subjects' environments. In this sense the present account differs importantly from other computational, or normative, explorations of psychiatric conditions. We do not argue that depression is an adaptive means to achieve a particular goal (Nesse, 2000; Stevens and Price, 2000). We do argue that depressive behaviour is due to a mismatch between the environment's characteristics and a subject's assumptions about them (a mismatch in parameters essentially, similar to previous work; Williams and Dayan 2005; Smith et al. 2004, 2005, 2006). However, our account shows how this may emerge from *normative* computations. This is of course a strong aetiological statement, and we will return to and refine it in chapter 6.

The immediate objective of the present chapter was to formalise the notion of control and analyse a few of its immediate sequels. We found that a prior over the extent of control in an environment alters the degree of branching of a decision tree. It has profound consequences on reinforcement processing. Simply stated, control renders the world nicer, more colourful and worth exploring. As rewards are exploitable and punishments avoidable, the world becomes nicer overall and actions vary more extensively in their rewarding properties. Only with control is it worth trying out new actions until the best possible action is found. Note again that

in the present formulation maximisation of rewards is equivalent to minimisation of punishments, and the effects of control apply equally to tasks defined in terms of rewards or in terms of punishment.

The chapter provides arguments for why a prior over control may useful in a more general RL setting. We have shown that expecting too much control leads to excessive exploration at the expense of exploitation. It remains to be seen whether a prior on control indeed provides computational advantages in larger RL problems (for example relative to E^3 (Kearns and Singh, 1998; Ghavamzadeh and Engel, 2007), or direct policy methods (Baxter and Bartlett, 2000)) and we have not shown how to combine such a prior with cached methods. Some possibilities are explored in related Bayesian reinforcement learning methods papers (Dearden et al., 1998, 1999; Friedman and Singer, 1999; Strens, 2000).

We would like to stress that the mathematical particularities of the present model are entirely arbitrary. For example, to assess the effect of entropy, we wrote outcome distributions as a mixture of a uniform and a delta function. Clearly, this is a very drastic reduction, and there may be better formulations that directly quantify the entropy. Future work will certainly examine the feasibility of using for example correlated Dirichlet processes to express the various notions of control in a more general, and hopefully in a mathematically more concise form.

Future work will also build on this formulation of control in an attempt to measure it directly in depressed populations. The formal definition can be used to define tasks that allow us to infer individuals' setting of their control priors and their willingness to modify and generalise it across environments. Finally, this can then be correlated with people's conscious assessments and should provide a way of relating the extensive work on perception of control to theoretical descriptions of affective decision making.

4.4.2 DOPAMINE

The attempt of a computational characterisations of psychiatric disorders was also motivated by the computational roles ascribed to particular neuromodulators (e.g. Montague et al. 1996; Dayan and Yu 2006; Niv et al. 2007). We saw that dopamine has relatively strong links to depression. Indeed, tonic dopamine appears to be the most natural neurobiological substrate for control. Mania is characterised by delusions of control, and is treated with DA antagonists. Increases in tonic DA increase specific motivational drives but also actions in general. Niv et al. (2005, 2007) give a detailed, quantitative account of a number of these effects by proposing that tonic DA reports the average reward expected from emitting actions per unit time. Such an opportunity cost notion is not unrelated to the formulation of controllably achievable reward here. As $\chi \to 1$, actions are increasingly worth the effort. Indeed, there are some indicators that tonic DA is not only enhanced by rewards, but also by controllable punishments (Cabib and Puglisi-Allegra, 1996; Horvitz, 2000). A litmus test of a link between control and dopamine would be to measure tonic DA levels in situations of uncontrollable rewards.

In terms of depression, it predicts a correlation between motivational deficits and prior expectations of no control. It appears to be the case that the most severely depressed patients suffer both from a motivational deficit and feelings of helplessness (Parker and Hadzi-Pavlovic,

1996), but specific tests are needed before this question can really be answered.

However, it is also clear that this is a vastly oversimplified picture. First, control is a complex construct of the goal-directed system, whereas dopamine is much more closely related to the habitual and motivational systems. It is difficult to see by what connections tonic DA levels could come to represent a value like controllably achievable reinforcement — indeed it is particularly difficult to see how it could do this independent of valence. Secondly, the consequences of using one and the same molecular substrate to represent such varied aspects of different affective systems are probably profound. Thirdly, pharmacological increases of tonic DA by amphetamine potentiates not only specific, but also general motivational drives, and this general motivational drive probably does not fit into our formulation of control.

4.4.3 Symmetry between rewards and punishments

There were two major motivations to analyse control in detail: first its prominent place in the thought of people researching depression. But secondly and much more specifically, we were motivated by the observation that exposure to uncontrollable reinforcers has effects that generalise across reinforcer value (figure 3.13; Goodkin 1976; Brickman et al. 1978; Overmier et al. 1980; Zacharko et al. 1983; Mineka and Hendersen 1985; Zacharko and Anisman 1991; Willner 1997; Gambarana et al. 1999; Gardner and Oswald 2001; Job 2002). This is a rather strong finding given the very different neurobiologial substrates of reward and punishment processing (see section 2.1.4). It is a prominent aspect of the experiments on LH that cannot be straightforwardly accounted for by a simple value-based system devoid of the notion of control (chapter 3) because no known link exists between analgesia (which is known to be inducible by shocks and stress), and decreased reward sensitivity (indeed, opioids tends towards the opposite effect). To account for the blunting symmetry seen in LH, our formulation of controllably achievable reinforcement is valence-free, in that is a measure only of the *normalised fraction* of the total reinforcement available in the environment (equation 4.4 and appendix B.3).

In the absence of experiments that directly assess goal-directed learning (such as reinforcer devaluation; Balleine and Dickinson 1998; Dickinson and Balleine 2002), in these models of depression, it appears that a behavioural insensitivity to reinforcers which is symmetrical in terms of valence is the strongest index for an involvement of control. In terms of human depression, the data is not strong enough to draw any conclusions. Some studies on the primary sensitivity to reinforcers (e.g. physiological responses to emotional scenes in movies; Rottenberg et al. 2002) have reported symmetrical effects, but these are not informative about the goal-directed system. Questionnaire data on the other hand seems to indicate a perceived hypersensitivity to punishments, in concord with a hyposensitivity to reinforcements (Lewinsohn et al., 1979), but this data is counfounded both by reports and by potential changes in primary sensitivity.

4.4.4 Human data on control

In terms of induction of helplessness, there are many implementations of the original animal LH paradigm in humans, but few of them are exactly interpretable in a rigorous reinforcement learning framework as used here. We do hope that the present formalisation will facilitate fur-

ther behavioural experimentation with human participants, both on the basic phenomenon of control inference and generalisation, and on its relationship to depression. Consider Miller and Seligman (1975): "master" students were exposed to stressful, loud noise which they could control, "yoked" students to the same noise but without control. Both were explicitly told that they could turn off the noise, and were given feedback as to whether it was their action that had turned off the noise, or whether the noise had stopped because it was scheduled to do so. All students were then given anagrams to solve. Master students were better at solving the anagrams than the yoked students. Hiroto and Seligman (1975) show that similar effects are apparent when students are first exposed to solvable/insolvable anagrams, and the effect also seems to occur after experience of uncontrollable rewards (Maier and Seligman 1976; Overmier et al. 1980; Job 2002 and Brickman et al. 1978; Mineka and Hendersen 1985 for anecdotal evidence in zoos, elderly care homes and after lottery wins).

At first sight, these are really quite impressive generalisations of the animal paradigm to human behaviour, but on the other hand they are also very loose generalisations (though see Costello 1978 for a critique). For one, the anagram task is a cognitive task not directly interpretable in terms of learning from the presented reinforcements. Then, rather than experienced lack of control, it is only perceived control that is relevant (Glass et al., 1973), and subjects have to be made to believe, quite explicitly, that failure on the task is indicative of their general ability (Roth and Kubal, 1975). Furthermore, some studies found better performance after helplessness induction, usually when only small amounts of helplessness induction were given (Roth and Bootzin, 1974; Wortman and Brehm, 1975; Mikulincer, 1988, 1994).

More relevant to our present concentration on animal data is that these experiments really do not distil out the various contributing factors. Is it that students who experienced failure were no more motivated to try other tasks? Or were they unable to use the feedback to infer the best action or simply insensitive to the positive feedbacks given? To our knowledge, most attempts to distinguish these causes in humans rely on verbal reports, and even those that do not are are at best ambiguous. For example, Alloy and Abramson (1982) give IS or ES, and then ask for judgements of control in a different tasks in which there is *no* control (unlike in the animal LH experiments). After ES they find accurate judgements, but after IS they see an illusion of control (rather than opposite). In terms of reward sensitivity, control over one's own shock exposure increases pain thresholds — an effect that is expected in the yoked, not the master subjects (Miller 1979, although see Badia et al. 1979 for complex effects of predictability). Unfortunately, we are not aware of any studies that attempt to dissociate these effects behaviourally, e.g. of studies that looked at appetitive or aversive learning after helplessness inductions. Indeed, because the learned helplessness theory as formulated by Maier and Seligman (1976) predicted a whole series of changes, including cognitive, motivational and emotional, such a dissection may not have seemed relevant.

Nevertheless, there are encouraging patterns in human data. We saw at a formal level that control affects reward exploitation, avoidance of punishment and exploration. Indeed, all of these might relate to standard notions of temperament, for example Cloninger (1987)'s tridimensional scheme with pleasure seeking, harm avoidance and novelty seeking. Decreasing levels of control here certainly would produce *both* decreased responsiveness to rewards and punishments, i.e. decreased reward seeking and decreased harm avoidance. In Cloninger's

scheme, depression is typically characterised as decreased pleasure seeking as opposed to alterations in harm avoidance (Otter et al., 1995; Ebstein et al., 2000; Compas et al., 2004; Myin-Germeys et al., 2003; Hettema et al., 2006b) or novelty seeking.

However, evidence from a variety of different fields question such a strong relationship: First, there is extensive comorbidity between depression and anxiety (Kessler, 1997; Mineka et al., 1998; Kaufman and Charney, 2000; Kendler et al., 2003b), and indeed DSM III placed anxiety and depression within one class. Second, depressed people do self-select themselves into high-risk environments and may by their own actions evoke negative feedback from others (Anisman and Matheson, 2005; Kendler et al., 1999, 2000). Third, we saw in chapter 2 that there is extensive evidence for a decreased sensitivity to punishments in depression. Fourth and maybe most related to the present discussion, Kendler et al. (2003a) look at the class of life events that preceding episodes of major depressive disorder and generalised anxiety disorder. They analyse a class of events characterised by "entrapment", which is, amongst those they analyse, most closely related to the present notion of control. They find that high ratings of entrapment most reliably leads to mixed as opposed to pure episodes of one disease or the other. If anxiety is taken as an increased sensitivity to punishments, then the notion of decreased control, as it is formalised here, leading to mixed depression and anxiety is not consistent. However, it is unclear that this notion of anxiety is necessarily correct (it may, for instance, be a better description of panic disorder).

Along similar lines the cognitive (Beck et al., 1979), LH (Maier and Seligman, 1976) and hopelessness theories (Abramson et al., 1989) all posit that a decreased perception of control is central to depression. One point we made in the literature review about the latter two theories is worth reiterating. It is generally found that depressed people attribute positive events to chance, and negative events to stable causes beyond their reach. This means that they cannot exploit positive or avoid negative events — precisely what is expected from a general lack of control. Importantly, the lack of control is applied without difference to both positive and negatively valenced events. However, once again, there is no direct behavioural evidence for alterations in exploration/exploitation strategies in depression.

The section on CMS (section 4.3.5.2) gives a possible replication of the internal *vs* external attributional dimension (Abramson et al., 1978) in addition to giving an interpretation to the variability in the sensitivity to CMS amongst individual rats (Strekalova et al., 2004). The control variable χ was either assumed to be shared across environments, or to be private to each environment. When encountering a new, unknown environment, subjects in the first case just applied the shared χ. In the latter case, subjects used a weighted sum of the χ variables, where the weights depended on the similarity between the new and the old environments. An internal attribution might correspond to the assumption that there is only one setting of χ across all environments. Control could then be said to be a feature of the agent, rather than the environments.

4.4.5 ANIMAL DATA ON CONTROL

We found that these simple formulations of control are capable of reproducing some effects seen in animal models of depression. They give qualitative accounts of the main effects seen

in learned helplessness, which have long been claimed to be related to control (Seligman and Maier, 1967; Seligman, 1975; Jackson et al., 1978). We have shortly presented a scheme to account for chronic mild stress, and other related models where animals are exposed to uncontrollable but mild stress (Willner et al., 1987; Cabib and Puglisi-Allegra, 1996).

What we have not yet addressed is that none of the stress-induced animal models are devoid of anxious effects: LH itself has been proposed to be a better model of post-traumatic stress disorder than depression (Maier and Watkins, 2005). Even a single exposure to a stressor can have lasting effects (Cordero et al., 2003; Mitra et al., 2005). In a very detailed, in-depth study, Strekalova et al. (2004) found that chronic mild stress produced anhedonia, either in combination with anxiety or not. Specifically, they gave animals several weeks worth of chronic mild stress, which produced a decrease in the time spent in a lit open box, decreased the time spent on the open fraction of an elevated O-maze and decreased the number of exits. In animals which did at the end display anhedonia (as measured by decreased preference of sucrose over water), they also found less exploratory behaviour (of a novel cage or a novel object). The finding that decreased reward sensitivity may go hand in hand with decreased exploratory behaviour is mirrored by our model. The induction of anxiety is more complex. It is also present in the animals that do not show anhedonia, and may thus simply be an unrelated process, but its presence in the animals with anhedonia goes against this interpretation.

Harding et al. (2004) provide probably the most unambiguous evidence that CMS has asymmetric effects on reward and punishment processing. They train rats to press a lever for reward with one tone, and to resist pressing the lever to avoid a punishment after a second tone. They find that CMS decreases the fraction of times rats press the lever for reward, but does not increase the fraction of times the rat do press the lever when a shock is predicted. If this asymmetry is real (and not, due e.g. to a floor effect), this cannot be accounted for by our theory. A similar asymmetry is proposed by Fontella et al. (2004). They take their starting point from the finding that both palatable and unpalatable food can affect the pain threshold of animals. Pleasant foods actually suppress tail-flick latency in normal animals. They find that, in normal rats, pleasant foods suppress TFL, and unpleasant foods do not affect TFL. In chronically stressed rats (1h daily immobilization for 40 days), pleasant foods have no more effect, and unpleasant foods produce a slight increase in TFL. They interpret this in terms of negative bias.

It may be that the chronic mild stress results in Pavlovian effects in addition to the instrumental control effects (see Dayan et al. (2006) and chapter 5), and it may be that these are large enough to offset the decreased sensitivity to punishments predicted by the decrease in perceived control. An interesting example is provided by Ghiglieri et al. (1997), who give animals appetitive training in a Y-maze. They simultaneously expose them to mild, chronic IS and find that this prevents the appetitive learning, as well as inducing an escape deficit. All of these deficits are sensitive to chronic antidepressant treatment. While this would be predicted by our formulation, an asymmetry appears when animals are trained before IS: IS did not disturb appetitive behaviours that had been learned previously, despite still inducing an escape deficit. Our present formulation cannot directly account for this. It may be that this effect is due to an interaction between Pavlovian and instrumental controllers, and there are good reasons to believe that these two controllers take precedence at different points in learning (Daw et al. 2005; Lengyel and Dayan 2007; see also chapter 5).

V

PAVLOVIAN INHIBITION

ABSTRACT

Depression is associated with the less efficient version of the serotonin reuptake mechanism. Yet, it is serotonin reuptake inhibitors (SSRIs) that are first-line drugs for depression. In healthy function, serotonin is thought to inhibit actions and report punishments, yet depression can be reliably re-induced by tryptophan depletion, which reduces serotonin levels. Here, we suggest that the combination of the two functions of aversive prediction and inhibition in one molecule have the computational effect of pruning a decision tree, i.e. preventing those decisions that have low expected outcomes. This has the overall effect of avoiding bad outcomes and results in higher average rewards. In the context of a highly simplified model of chains of affectively-charged thoughts, we show how a drop in this inhibition results in unexpectedly large negative prediction errors and a large aversive shift in the reinforcement statistics.

5.1 INTRODUCTION

The previous two chapter have concentrated on computational analyses of behavioural phenomena, and the neuromodulatory basis of these was in the relative background. In contrast, we here concentrate on the most prominently involved neuromodulatory aspect of depression — serotonin. We have previously reviewed evidence for its involvement in depression and animal models of depression, but also anxiety and other disorders (section 2.1.4). In this chapter,

we suggest that three apparently separate or even contradictory facts and hypotheses about 5-HT are actually linked. The first is perhaps the main functional association made for 5-HT, that it is involved in the prediction of aversive events, as a form of opponent (Solomon and Corbit, 1974; Dickinson and Dearing, 1979; Dickinson and Balleine, 2002) to dopamine (Carter and Pycock, 1978; Costall et al., 1979; Deakin, 1983; Deakin and Graeff, 1991; Fletcher, 1996; Kapur and Remington, 1996; Daw et al., 2002; Esposito, 2006). The second fact is that serotonin is involved in forms of behavioural inhibition (Soubrié, 1986; Gray, 1991; Schmajuk et al., 1996), preventing or curtailing ongoing actions. The third issue is the collection of psychopharmacological findings implicating 5-HT in animal models of depression and anxiety (Willner, 1985b; Graeff et al., 1998; Maier and Watkins, 2005), and notably that depleting 5-HT (by dietary depletion of its precursor, tryptophan) in human subjects who have recovered from depression, reinstates an acute, temporary, but fulminant re-experience of subjective phenomena of depression, as assessed by various rating scales (Young et al., 1985; Delgado et al., 1990; Moreno et al., 1999; Smith et al., 1999). The second fact seems orthogonal to the first and third, which are themselves in apparent contradiction. If 5-HT is really involved in predicting aversive outcomes, *depleting* it should, if anything, have positive rather than negative affective consequences.

We suggest that the missing link comes from considering the influence that Pavlovian predictions have over ongoing behaviour. This is straightforwardly seen in conditioned suppression (Estes and Skinner, 1941), a standard workhorse test for aversive predictions some aspects of which were explored in chapter 3. Subjects are trained instrumentally to press a lever to get access to reward, and classically about the predictive relationship between a light and a shock. If, whilst they are pressing the lever, the light is turned on, they will tend to reduce or suppress their lever-pressing. Neither the theoretical nor the neurobiological status of this interaction is completely resolved, though there is some evidence of the involvement of 5-HT in the nucleus accumbens in its realization (Fletcher, 1995; Fletcher and Korth, 1999; Graeff, 2002). In chapter 3, we concentrated on how the reinforcement history and availability of a blunting action can interfere with the acquisition of this response. By contrast, here we analyse the consequences of the fact that serotonin is involved in several computational components of the behaviour at once.

We treat a subset of the inhibitory processes associated with Gray's behavioural Inhibition System (Gray, 1991; Deakin and Graeff, 1991; Gray and McNaughton, 2003; McNaughton and Corr, 2004) in terms of a Pavlovian 'action' that is specified over the course of evolution as being a pre-programmed response (Blanchard and Blanchard, 1988) to a potentially dangerous situation. This response is assumed to be available in a reflexive manner without the requirement for any further learning, and which is appropriate in many, though not all (Breland and Breland, 1961; Dayan et al., 2006), circumstances. This directly ties the first two issues above together, with inhibition arising from the aversive prediction. It effectively prunes (Knuth and Moore, 1975; Baum and Smith, 1997) the decision tree and as such leads to a critical *bias* in the interaction between subjects and their environments. This bias is towards optimism, since states and actions with potentially negative consequences are incorrectly (over)valued and underexplored because of the inhibition. When inhibition fails, though, there are two adverse consequences. First, the inhibition is no longer a crutch for instrumental action choice — so subjects would have to learn to avoid potentially bad situations rather than being able to rely

on this intrinsic mechanism. Second, characteristic inconsistencies between the predicted and actual values arise, with the actual values encountered being more negative than predicted, though also actually more realistic.

To explore the marked consequences for affective evaluations of direct inhibition of action, together with the repercussions when 5-HT is compromised, we build a highly simplified model of trains of thought. In this treatment, which is not intended to be physiologically truthful in detail, we consider thoughts as actions which lead from one state of belief to the next. Trains of thought gain worth by virtue of a group of terminal states being preassigned either positive or negative affective values. 5-HT reports on a particular aspect of the expected aversive consequence of thinking a thought, and directly inhibits thoughts predicted to lead towards negative terminal states. Under normal circumstances, the ultimate effect of serotonergic inhibition is to bias evaluations to be unduly optimistic, as above. Depletion of 5-HT leads to more realistic (and thus inevitably more pessimistic) predictions. Boosting 5-HT again restores the *status quo*. Of course, this highly simplified model cannot possibly, by itself, accommodate all the diverse and confusing roles of 5-HT. Further, we focus on the consequences of inhibition for quantities associated with information processing such as predicted values and chosen actions. Nevertheless, engagement of the behavioural inhibition system has been equated (Gray and McNaughton, 2003) with anxiety, and so we speculate in the discussion on a possible link with aspects of anxiety and depression.

The next section defines the model of trains of thought more formally. Section 3 considers normal, biased, learning, and the consequences of impairments to 5-HT processing. We save for section 4 a broader discussion of data and theories pertaining to 5-HT.

5.2 METHODS

5.2.1 THE MODEL

Figure 5.1 illustrates our underlying model of trains of thought. It is highly simplified, and does not stand as a faithful rendition of any psychological model of thinking. However, it allows us to focus directly on a role for 5-HT in behavioural inhibition.

The model contains terminal states ($\mathcal{O}_+, \mathcal{O}_-$), which are arbitrarily assigned positive and negative affective values respectively, and internal states ($\mathcal{I}_+, \mathcal{I}_-$) which are preferentially, though sparsely, connected with their own 'sign' of internal and terminal states. A thought is modelled as a transition between states along extant connections; a train of thoughts ends up in one of the terminal states. In this simple model, the value of an internal state is the average value of the terminal states to which it ultimately leads. By biasing the selection of actions (*ie* thoughts), 5-HT biases the evaluation of states and the experience of positive and negative affective outcomes.

More formally, the model is a form of Markov decision process (see Sutton and Barto 1998), with four sets of sparsely interconnected states $\{\mathcal{I}_\pm, \mathcal{O}_\pm\}$. Two sets, $\mathcal{O}_+$ and $\mathcal{O}_-$ (each with 100 elements in the simulation) are associated respectively with positive ($r(s) \geq 0, s \in \mathcal{O}_+$) and

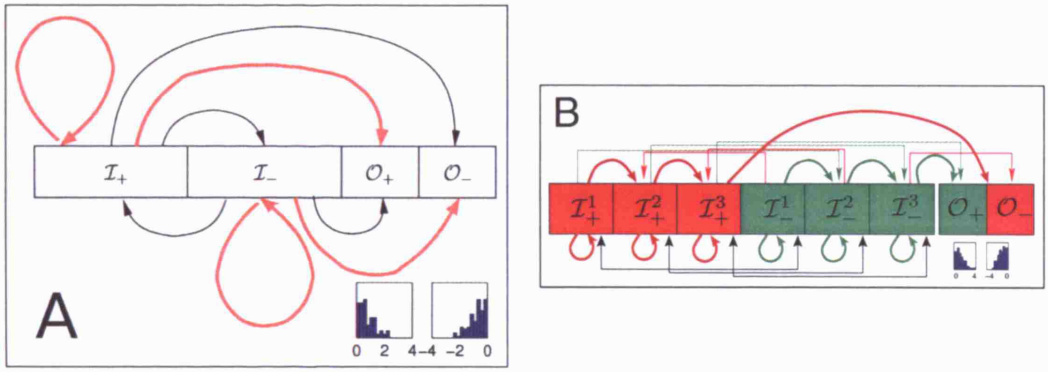


FIGURE 5.1: Markov models of thought. **A**: The abstract state space is divided into the 4 blocks shown. The right two, $\mathcal{O}_+$ and $\mathcal{O}_-$, are associated with direct affective values $r(s)$ (inset histograms); the left two, $\mathcal{I}_-$ and $\mathcal{I}_+$, are internal. Actions (thoughts) move from one state to another, in a sort of sparse trellis. States in each internal block $\mathcal{I}_+$ and $\mathcal{I}_-$ preferentially connect with each other and their respective outcome states $\mathcal{O}_+$ and $\mathcal{O}_-$. However, each state has links to states in the other block. There is an approximate balance of positive and negative affect in the model as a whole. **B**: Similar state space to A, but with a more explicitly deep structure. State in $\mathcal{I}_+^1$ mainly lead to $\mathcal{I}_+^2$, or back to themselves. The last states in each of the two chains (here $\mathcal{I}_+^3$ and $\mathcal{I}_-^3$) always preferentially lead to the outcome state $\mathcal{O}_+$ and $\mathcal{O}_-$.

negative affective values ($r(s) \leq 0, s \in \mathcal{O}_-$; both drawn from suitably truncated 0-mean, unit variance, Gaussian distributions, see inset histograms in figure 5.1) and are terminal states. The other sets, $\mathcal{I}_+$ and $\mathcal{I}_-$ (each with 400 elements) contain internal states and are associated with 0 affective values ($r(s) = 0$).

Each element of $\mathcal{I}_+$ has 8 outgoing connections, 3 to other (randomly chosen) elements in $\mathcal{I}_+$; 3 to randomly chosen elements in $\mathcal{O}_+$; and 1 each to randomly chosen elements in $\mathcal{I}_-$ and $\mathcal{O}_-$. Similarly, each element of $\mathcal{I}_-$ has 8 outgoing connections, 3 to other (randomly chosen) elements in $\mathcal{I}_-$; 3 to randomly chosen elements in $\mathcal{O}_-$; and 1 each to randomly chosen elements in $\mathcal{I}_+$ and $\mathcal{O}_+$. Thoughts are modelled as actions a following these connections, labelled by the identities of the states to which they lead.

To look at effects of impulsivity, we will then subdivide the states $\mathcal{I}_+$ and $\mathcal{I}_-$ further into say K compartments (figure 5.1 shows this for $K = 3$). $\mathcal{I}_+^K$ will be equivalent to the $\mathcal{I}_+$ of figure 5.1, and for all other compartments $k < K$, the role of $\mathcal{O}_+$ is replaced by $\mathcal{I}_+^{k+1}$. The same applies to the negatively valenced states.

The dynamics of the world are that a train of thought consists of starting from a random state (for this section, chosen equally across $\mathcal{I}$), and following successive transitions until either the train runs into a terminating state or is inhibited via a serotonergically-controlled mechanism (see below).

A policy $\pi_s(a)$ is a probability distribution over possible thoughts a at state s (and thus over possible next states s' given state s). Dynamic programming (Sutton & Barto, 1998) leads to a

value function $V^\pi(s)$ over states s, defined by $V^\pi(s) = r(s), s \in \mathcal{O}_\pm$, and

$$V^\pi(s) = \gamma \sum_a \pi_s(a) V^\pi(a) = \gamma \sum_{s'} \pi_s(s') V^\pi(s') \tag{5.1}$$

where γ is a discount factor ($\gamma = 0.9$ in our simulations), and a $\mathcal{Q}^\pi(s,a)$ function over states and thoughts defined for those actions that exist by

$$\mathcal{Q}^\pi(s,a) = \gamma V^\pi(a) \tag{5.2}$$

There are also optimal value $V^*(s)$ and $Q^*(s,a)$ functions, which are associated with any policy $\pi_s^*(a)$ that maximizes the long-run affective consequences of the train.

Since we are primarily interested in classical conditioning, we start by considering a base policy in which the probability of each possible thought is equal (1/8 for the internal states). This defines a set of normative values for all the states, with the values for $\mathcal{I}_+$ being, on average, greater than those for $\mathcal{I}_-$. The structure of a train is that it starts in a state in $\mathcal{I}_\pm$, bounces around states in $\mathcal{I}_\pm$ for some number of thoughts, and ultimately terminates in a state in $\mathcal{O}_\pm$. We might very crudely consider the relative proportion of states in $\mathcal{I}_-$ compared with those in $\mathcal{I}_+$ as a form of negative rumination, since these are states generally associated with negative values, and more likely terminate in the actually aversive class $\mathcal{O}_-$. The present setup is symmetric though.

5.2.2 SEROTONIN

There are various ways to model the precise effect of serotonin and dopamine on the base policy. We focus exclusively on 5-HT, and consider the simplest possibility associated with the suggestion of (Daw et al., 2002), that 5-HT represents negative values of states, and that it can stochastically terminate aversive trains of thought, with a probability of *continuation* of

$$p^\pi_{\text{5-HT}}(s) = \min\left(1, \exp\left(\alpha_{\text{5-HT}} V^\pi(s)\right)\right) \tag{5.3}$$

where $\alpha_{\text{5-HT}}$ is a multiplicative factor that scales the impact of the 5-HT. The more disastrous the potential sequelæ of state s, the more negative $V^\pi(s)$, and so the less likely the thought is to be continued. On the other hand, even slightly positive values will essentially veto any termination. This introduces an asymmetry into the model. Other possibilities for the information reported by 5-HT are considered in the discussion.

5.2.3 LEARNING

We use temporal difference learning (TD; Sutton and Barto (1998)) to acquire the values of states under the simplest policy defined by the combination of equal change for each thought together with serotonergic inhibition. TD specifies an online learning rule which, if the subject

takes action a at state s, then the change in the estimated value is

$$\Delta V^{\pi}(s) = \epsilon \begin{cases} 0 & \text{if the train is inhibited} \\ r(a) + \gamma V^{\pi}(a) - V^{\pi}(s) & \text{otherwise} \end{cases} \tag{5.4}$$

where ϵ is a learning rate and remember that a defines the next state s' deterministically. That $V^{\pi}(s)$ does not change given termination implies that learning is only slowed for these states, rather than being biased towards 0. However, the qualitative characteristics of our results would not be changed if instead

$$\Delta V^{\pi}(s) = \epsilon \begin{cases} -V^{\pi}(s) & \text{if the train is inhibited} \\ r(a) + \gamma V^{\pi}(a) - V^{\pi}(s) & \text{otherwise} \end{cases} \tag{5.5}$$

which would be the more conventional application of TD learning in this context.

In the results, we show values after substantial learning (20000 trains); plus the consequences of manipulating serotonin (by manipulating $\alpha_{5\text{-HT}}$) once the values are already acquired. We also calculate the true values $V_{\text{true}}(s)$ for states under the base policy using methods from dynamic programming (Sutton and Barto, 1998).

5.3 Results

5.3.1 Behavioural inhibition

Figure 5.2 shows the consequence that behavioural inhibition through learning has on the estimated values of states. Figure 5.2A shows that for the base policy, 20000 learning steps are ample to acquire a reasonable values $V_{\text{est}}(s)$ for the states (the remaining discrepancies from $V_{\text{true}}(s)$ arise from the stochasticity in the choice of action together with the fixed learning rate). By comparison, figure 5.2D shows that setting a large value of $\alpha_{5\text{-HT}} = 20$ biases learning significantly, with the result that low valued states are much less well visited and explored. Of course, the extent to which this is true depends on the initial values for the states (all of which are set to 0 in the simulation). Figure 5.2E shows how frequently each of the outcome states was reached in a run (as a function of its outcome $r(s)$). Since behavioural inhibition terminates trains on their way to potential disaster, aversive terminal states are sampled less (shown by the red regression line), which is consistent with the bias of the estimated value. Figures 5.2C;F show these effects as a function of $\alpha_{5\text{-HT}}$. The greater the inhibition the worse estimated are the values (C), particularly for aversive states; but the more benign is the exploration (F). Learning with inhibition leads to an optimistic set of values. However, this is coupled with a more aggressive rejection of all actions even mildly associated with negative outcomes.

5.3.2 Serotonin depletion

Given the values $V_{\text{est}}^{\alpha_{5\text{-HT}}}(s)$ learned under $\alpha_{5\text{-HT}} = 20$, the steady-state transitions probabilities can be calculated for any new $\alpha_{5\text{-HT}} \neq 20$ simply by working out the probability of inhibition for

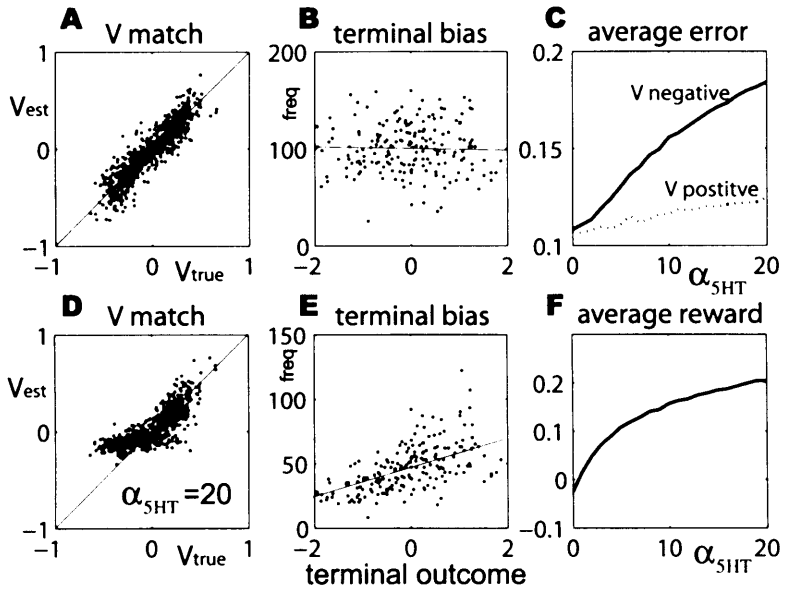


FIGURE 5.2: Learning with behavioural inhibition. (A;B) with $\alpha_{5\text{-HT}} = 0$, for one particular learning run, the values V_{est} match their true values V_{true} under an equal-sampling exploration policy (A); and trains of thought end in terminal states $\mathcal{O}_-, \mathcal{O}_+$ equally often as a function of their actual outcomes (B; the red line is the regression line). (D;E) with $\alpha_{5\text{-HT}} = 20$, negative V values are poorly estimated (since exploration is progressively inhibited for larger $\alpha_{5\text{-HT}}$), and the more negative the value of the outcome, the less frequently that outcome gets visited over learning (E). Importantly, there is an optimistic *under*estimate of the negative value of state. (C) shows the root mean square error (averaging over 20 runs) for states with positive (dotted) and negative V_{true} values as a function of $\alpha_{5\text{-HT}}$. The effect of the sampling bias is strikingly apparent, preventing accurate estimates mainly of the negatively valued states. (F) shows the average reward received during learning as a function of $\alpha_{5\text{-HT}}$ — the benefits of behavioural inhibition are apparent.

each state. From the results in the previous paragraph it is apparent, that the negative values of states are underestimated. Thus, a computation of the termination probabilities based on these values and a lower value of $\alpha_{5\text{-HT}}$ is expected to lead to higher transition probabilities into states with negative outcomes. The new transition probabilities define a new policy, and we can thus compute the value under this new policy and compare it to that under $\alpha_{5\text{-HT}} = 20$. Two statistics from the process are particularly significant. One is just the average affective outcome (the average value) of trains of thought in the model. The second is a measure of the surprise at each outcome, measured by the prediction error

$$\delta = r(a) - V_{\text{est}}^{\alpha_{5\text{-HT}}}(s) \tag{5.6}$$

for the last transition of a chain from state $s \in \mathcal{I}_{\pm}$ to a state $a \in \mathcal{O}_{\pm}$. We may expect negative prediction errors

$$\delta_{-} = \begin{cases} \delta & \text{if } \delta < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

to be of special importance, because of substantial evidence that aversive outcomes whose magnitudes and timing are expected so they can be prepared for, have substantially less disutility than outcomes that are more aversive than expected (at least for physiological pains, see Rachman and Arntz (1991)).

Figure 5.3 shows the consequences of learning under full inhibition and then wandering through state space with reduced inhibition. The change in the average terminal affective value as a percentage of the case during learning that $\alpha_{5\text{-HT}} = 20$ is shown in figure 5.3A. As was already apparent in figure 5.2F (which averaged over the whole course of learning), large costs are incurred for large reductions in inhibition. For $\alpha_{5\text{-HT}} = 0$, the average reward is actually negative, which is why the curve dips below -100%. This value is relevant, since the internal environment is approximately symmetric in terms of the appetitive and aversive outcomes it affords. Subjects normally experience an *optimistic* or rosy view of it, by terminating any unfortunate trains of thought (indeed 55% of their state occupancy is in $\mathcal{I}_{+}$ compared with $\mathcal{I}_{-}$). Under reduced 5-HT, subjects see it more the way it really is (the ratio becomes 50%).

Figure 5.3B;C show comparative scatter plots of the terminal prediction errors. Here, we consider just the last transition from an internal state to an outcome state. Prediction errors here that are large and negative, with substantially more aversive outcomes than expected may be particularly damaging. Figure 5.3C compares the average terminal prediction errors for all transitions into states in $\mathcal{O}_{-}$ with no serotonergic inhibition $\alpha_{5\text{-HT}} = 0$, to those for the value $\alpha_{5\text{-HT}} = 20$ that was used during learning. For the case that $\alpha_{5\text{-HT}} = 20$, the negative prediction errors are on average very small (partly since the probability of receiving one is very low). With reduced inhibition, the errors become dramatically larger, potentially leading to enhanced global aversion. By comparison, as one might expect, the positive prediction errors resulting from transitions into $\mathcal{O}_{+}$ are not greatly affected by the inhibition (figure 5.3B).

5.3.3 Recall bias

Three additional effects enrich this admittedly partial picture. One, which plays a particularly important role in the cognitive behavioural therapy literature, is that depressed patients have a tendency to prefer to *recall* aversive states or memories (Blaney, 1986; Klaassen et al., 2002). A simple way to model that is to bias the start distribution for sampling, favouring states $s \in \mathcal{I}$ with *lower* values $V(s)$ (this favours choices in $\mathcal{I}_{-}$ over $\mathcal{I}_{+}$). Figure 5.4A shows the consequence of doing this according to a simple softmax $p_{\text{start}}(s) \propto \exp(\beta V(s))$. These curves, as in figure 5.3A show the percentage average utility compared with $\alpha_{5\text{-HT}} = 20, \beta = 0$ across values of $\alpha_{5\text{-HT}}$, and for $\beta = -10, -9, \ldots, 10$. As might be expected, biasing the starting point to $\mathcal{I}_{-}$, and, even worse, to those particular states in $\mathcal{I}_{-}$ that are most deleterious, has a big negative impact on average utility. For $\alpha_{5\text{-HT}} = 0; \beta = -10$, occupancy of $\mathcal{I}_{+}$ relative to $\mathcal{I}_{-}$ becomes a paltry 27% as subjects ruminate (Nolen-Hoeksema, 1991; Smith et al., 1997) negatively.

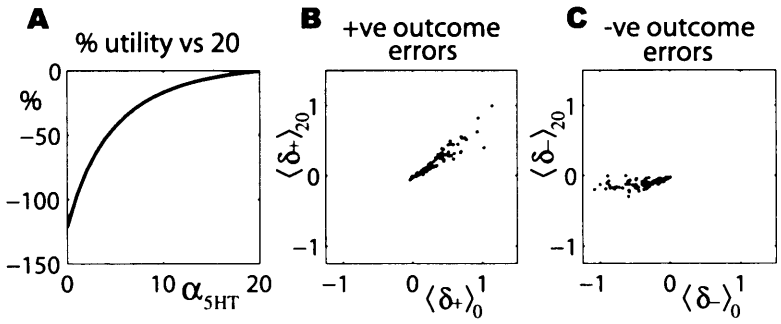


FIGURE 5.3: Reduced inhibition. These graphs show statistics of the effect of learning V values with $\alpha_{5\text{-HT}} = 20$, and then suffering from reduced serotonin $\alpha_{5\text{-HT}} < 20$ during sampling of thoughts. For a given thought environment, these are calculated in closed form, without estimation error. A) as is also evident in figure 5.2F, the average affective return is greatly reduced from the value with $\alpha_{5\text{-HT}} = 20$, in fact for the extreme value of $\alpha_{5\text{-HT}} = 0$, it becomes slightly negative (reflecting a small sample bias in the particular collection of outcomes). B;C) normalized *outcome* prediction errors at the time of transition to $\mathcal{O}_+$ (B) or $\mathcal{O}_-$ (C) for $\alpha_{5\text{-HT}} = 20$ against $\alpha_{5\text{-HT}} = 0$. These reflect the individual probability that each terminal transition goes to $r(s)$ from $V(s')$ for $s \in \mathcal{O}$ and $s' \in \mathcal{I}$, including all the probabilistic contingencies of termination, *etc.* They are normalized for the two values of $\alpha_{5\text{-HT}}$. Terminations in $\mathcal{O}_+$ are largely unaffected by the change in inhibition; terminations in $\mathcal{O}_-$ with negative consequences, have greatly increased negative prediction error.

5.3.4 REWARD SEEKING

The second factor is our restriction to just inhibition of trains of thought rather than a more finescale manipulation of the relative probabilities of different thoughts. The action values $\mathcal{Q}(s,a)$ can be used to choose amongst the available next states in a way that is guided by their Pavlovian values, *ie* Pavlovian withdrawal from, or instrumental deselection of, actions a associated with negative $\mathcal{Q}(s,a)$ values is as possible as is approach to or choice of actions a associated with positive $\mathcal{Q}(s,a)$ values. Such a control can be incorporated in a straightforward manner by choosing action a in state s according to a softmax $\pi_s(a) \propto \exp(\gamma \mathcal{Q}(s,a))$, where γ controls the degree of influence of the $\mathcal{Q}$ value. Figure 5.4B shows the effects of setting γ to be negative (as if subjects *prefer* transitions leading to aversive outcomes) or positive. It is apparent that rather extreme values of γ can also significantly aggravate or suppress the effect of $\alpha_{5\text{-HT}}$. For the highest positive values of γ the curves reverse shape, showing that it can be beneficial *not* to inhibit trains of thought. This arises since the model of figure 5.1 was chosen to have the extreme property that there is always the possibility of avoidance (in that all the states in $\mathcal{I}_-$ admit at least one action that leads to $\mathcal{I}_+$), and inhibiting trains of thought removes this outcome.

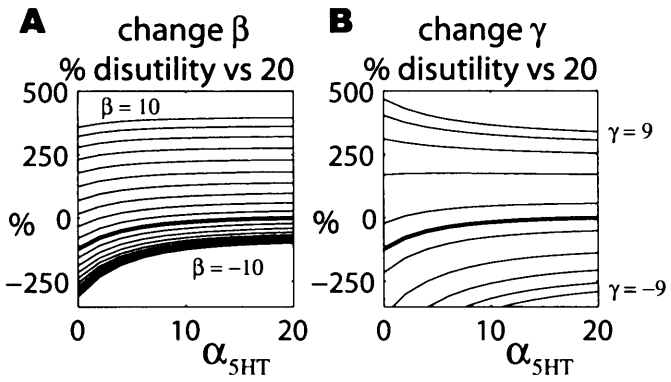


FIGURE 5.4: Both plots are in the same form as figure 5.3A, showing the percentage utilities compared with the standard learning case $\alpha_{5\text{-HT}} = 20$, as a function of $\alpha_{5\text{-HT}}$ (the emboldened blue curve is exactly that in figure 5.3A). A) Given a mood-dependent bias on the starting state, with $p_{\text{start}}(s) \propto \exp(\beta V(s))$, the plots show the consequences of various values of β. Negative β, favoring low value states, leads to substantially negative average outcomes. B) Instrumental control of action choice, a putative model of dopaminergic effects, can also either exacerbate or improve the outcomes, depending on the value of the parameter ϕ governing a softmax choice of actions.

5.3.5 IMPULSIVITY

Finally, it may be that the present implementation of inhibition could lead to short-sighted actions in deeper environments. Consider an environment, where all large rewards are hidden behind small punishments, and vice-versa. Large $\alpha_{5\text{-HT}}$ may prevent visits to the states associated with small punishments and thus prevent visits to the states behind them that are associated with large rewards. This effect is closely related to impulsivity, which itself has been associated with an impaired behavioural inhibitory system (Deakin, 2003). We first show that all the conclusions up to now also apply to an environment of depth K in which the reward structure is still bipartite, i.e. in which no good states are primarily accessible via negative states. Let states $\mathcal{I}_+^k$ of positive valence and at level $1 \leq k \leq K-1$ preferentially lead to states in the same level and valence $\mathcal{I}_+^k$ (with probability 3/8) or to states $\mathcal{I}_+^{k+1}$ (with probability 3/8) of the same valence, but one level closer to the outcomes $\mathcal{O}$. With smaller probability (1/8) they can also lead to states of the opposite valence at the same level $\mathcal{I}_-^k$ or one level up $\mathcal{I}_-^{k+1}$. States $\mathcal{I}^{K-1}$ have connections as shown in figure 5.1B and are the only states that can lead to outcomes $\mathcal{O}$. Figure 5.5A shows the true values of states without inhibition and their estimated values with inhibition. There is a clear positive bias for all negatively valued states. Figure 5.5B shows that the outcomes are still more frequently positive than negative, and figure 5.5C shows the effect of altering γ. As in figure 5.4B, negative γ lead to preferential selection of actions leading to states with negative values, and the overall average value is increased by increasing γ.

However, the situation changes if the states with large positive outcomes are primarily accessible through punished states. The dash-dotted line in figure 5.5D shows that the states $\mathcal{I}_+^K$ at the final level K before the outcomes are now punished, while the states $\mathcal{I}_-^K$ are rewarded. The values of states $\mathcal{I}_+^3$ is now more negative than their counterpart without inhibition. This

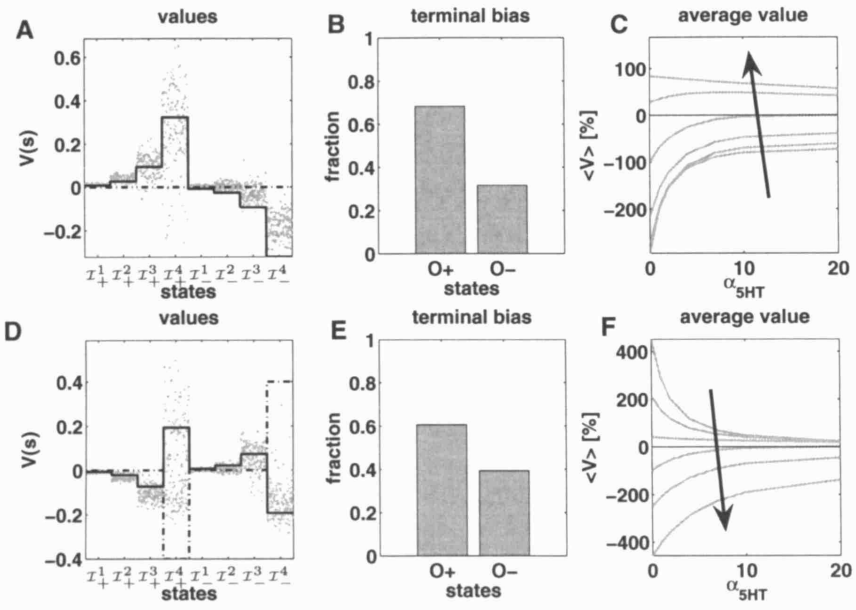


FIGURE 5.5: Inhibition in a deep environment. The outcomes $\mathcal{O}$ are approached by sequentially walking through $K = 4$ levels. Only $\mathcal{I}^4$ states lead to outcomes. (A,D): True values without inhibition are shown by black line. It is constant for each level and valence as, or illustration, all outcomes were assigned the same positive value (+1 or -1). The reward of the states $\mathcal{I}$ is zero and shown by the dash-dotted line. The grey point display the estimated values of the states under inhibition $\alpha_{5\text{-HT}} = 20$. There is a positive bias in all states, but it is more pronounced in the states with true negative values. In (D), the dash-dotted line indicates that states $\mathcal{I}_+^4$ now carry reward -0.4, while states $\mathcal{I}_-^4$ carry reward $+0.4$. States $\mathcal{I}_+^k$ for $k = \{1, 2, 3\}$ now have true negative values and $\mathcal{I}_-^k$ for $k = \{1, 2, 3\}$ have true positive values. (B,E): Probabilities of ending thought sequence in $\mathcal{O}_+$ or $\mathcal{O}_-$. (C,F): Effect of preferentially choosing actions according to their valence on the average value of states. The arrow indicates increasing γ. In (C), larger γ are advantageous, in (F), smaller γ are better.

is because thoughts are often interrupted before they can proceed through (the punished) $\mathcal{I}_+^4$ to (the rewarded) $\mathcal{O}_+$. Overall, the states in $\mathcal{I}_+$ are now predominantly negative, and those in $\mathcal{I}_-$ predominantly positive. However, inhibition does still lead to an overall positive outcome bias (figure 5.5E), and diminishing $\alpha_{5\text{-HT}}$ with $\gamma = 0$ is still unfavourable (figure 5.5F, black line). However, the effect of γ is now reversed (see arrow in figure 5.5F): Negative γ will now predominantly lead to choices in $\mathcal{I}_+$, and thus to choices with longer-term positive outcomes (as there is no more inhibition). Thus, when $\alpha_{5\text{-HT}}$ is lowered in an environment in which rewards lurk behind punishments, choosing negatively valued actions may actually be advantageous.

5.4 DISCUSSION

In this chapter, we studied a very simple Markov decision process model of affectively-charged thoughts, and showed various aspects of the influence of behavioural inhibition on the experience of appetitive and aversive outcomes, predictions and prediction errors. The model formalises behavioural inhibition as a Pavlovian control process that arrests internally-directed thoughts (and likewise externally-directed actions) that are predicted to lead to aversive consequences. Overall this is favourable, leads to enhanced average rewards, and is related to pruning (Knuth and Moore, 1975; Baum and Smith, 1997). However, the consequences can also be deleterious (Breland and Breland, 1961; Dayan et al., 2006). Compromising inhibition in the model has two related consequences. First, the values of states are revealed to be overly optimistic. Second, control is disturbed, with aversive chains being insufficiently deselected.

5.4.1 BEHAVIOURAL INHIBITION SYSTEM

We suggested that this form of behavioural inhibition arises through predictions of aversive outcomes, tied to serotonin's putative role in reporting aversive prediction errors as an opponent to dopamine. This comes directly from the original notions of behavioural inhibition and serotonergic effects from Gray, Deakin, Graeff and their colleagues (Gray, 1991; Deakin, 1983; Deakin and Graeff, 1991; Gray and McNaughton, 2003; McNaughton and Corr, 2004); however, it is perhaps best seen as a subset of the current version of Gray's Behavioural Inhibition System (BIS; Gray and McNaughton 2003). One salient difference is that BIS is suggested as being primarily engaged by *conflict*, rather than ongoing predictions of future aversive outcomes. Of course, a main source of conflict is that between approach and avoidance, with the latter coming from these aversive predictions. An interesting consequence of dividing the prediction of the value of future outcomes between two separate opponent systems is that it is indeed possible to have simultaneous appetitive and aversive expectations, as opposed to just one combined, net, prediction. Although in this chapter, we used the net prediction to control inhibition, it would be interesting to explore other possibilities associated with the BIS view, such as that any aversive prediction could arrest ongoing action, even if outweighed by appetitive predictions.

Further, rather than have the aversive predictive *values* of states lead to termination of trains of thought, it is possible that the negative prediction error (δ_- from equation 5.7), which (Daw et al., 2002) suggested is being reported by phasic serotonin, could be responsible instead, or alternatively, more like the tonic dopaminergic signal that Niv et al. (2005, 2007) postulated to report average reward and energise behaviour, that a more tonic serotonergic signal averaging aversion over longer time horizons, could be responsible.

Another difference between our account and the full BIS is that, in the latter, although actions are indeed inhibited in the face of conflict, the BIS is then suggested as initiating a set of behaviours (such as exploration or risk assessment) to resolve that conflict. The set of preparatory Pavlovian actions associated with aversive predictions appears to be more refined than that associated with appetitive predictions (mostly just approach), with a wide range of different defensive possibilities being selected between according to the nature and proximity of the

threat (Blanchard and Blanchard, 1988; McNaughton and Corr, 2004). One class of these is even laid out along columns of the peri-acqueductal gray (PAG; Bandler and Shipley 1994). Nevertheless, any of these defensive manoeuvres would interrupt the ongoing chain of actions, and this is what we modelled. Risk assessment and exploration are of most obvious use in the face of uncertainty and ignorance, whereas conditioned suppression, and thus the sort of inhibition that we consider, remains even after substantial learning. It would certainly be worth going one stage further, modelling the interruption in terms of a switch between different Markov decision problems, with new information changing the transition and payoff structures.

5.4.2 TRYPTOPHAN DEPLETION

Given that we see the effect of TrD as an acute reduction in $\alpha_{5\text{-HT}}$ after learning with elevated $\alpha_{5\text{-HT}}$ has taken place, *ie* as a *de*crease in behavioural inhibition of actions leading to negative states, the effects of acute tryptophan depletion (TrD; Bell et al. 2001) studies, in which CNS levels of serotonin are reduced by up to 90% in human or animal subjects (see section 2.1.4) are of course of particular relevance to our model. Although the particular behavioural chains analysed in this chapter have not been the subject of experimental scrutiny, there is by now a considerable body of literature on the effects of TrD on normal human functioning. In broad agreement with the results from this chapter, various effects have been related to decreased reward processing (Murphy et al., 2002; Klaassen et al., 2002; Roiser et al., 2006), decreased behavioral inhibition (LeMarquand et al., 1999; Bjork et al., 2000; Deakin, 2003; Anderson et al., 2003; Schweighofer et al., 2006, 2007), rumination (Smith et al., 1999), and, more indirectly and contentiously, increased aggressiveness (Bjork et al., 2000; Walsh and Dinan, 2001). In further agreement, tryptophan depletion does have greater effects on subjects who have the less efficient version of the serotonin reuptake mechanism (Lesch et al., 1996; Neumeister et al., 2002; Roiser et al., 2006; Hariri and Holmes, 2006), i.e. in subjects that are putatively exposed to higher levels of average serotonin throughout development (potentially extenuated by adaptive processes; Hariri and Holmes 2006).

It is even better-known that TrD produces a severe, dose-dependent relapse of depressive symptomatology in formerly depressed patients (section 2.1.4; Young et al. 1985; Delgado et al. 1990; Smith et al. 1997; Moreno et al. 1999), or in patients with risk factors such as a family history of depression or one version of the short 5HTTLPR allele (Neumeister et al., 2002). There is as yet no such positive result for anxiety, although panic disorders are aggravated by TrD (Klaassen et al., 1998; Miller et al., 2000).

The most obvious predictions from this model come from manipulations of 5-HT. In particular, it would be most interesting to train and test subjects with or without tryptophan depletion on a Markov decision problem of the type we discussed, and study their exploration and exploitation behavior, both of which we would expect to be affected. This could use external, observable, actions; it would also be interesting to seek measures of the execution of affective trains of thought, and study their perturbation under serotonergic manipulation. In designing such studies, it is important to bear in mind the potentially opponent instrumental and Pavlovian effects, in just the same way that boosting dopamine and monitoring the effects on negative automaintenance may be confusing. Note that although there are most interesting

data on TrD in simple probabilistic and delay-discounting tasks (Anderson et al., 2003; Rogers et al., 1999; Mobini et al., 2000a,b; Murphy et al., 2002; Rogers et al., 2003; Cools et al., 2005; Roiser et al., 2006), these studies do not encompass the sorts of behavioral chains that we propose 5-HT to be able to halt.

5.4.3 SEROTONIN AND DOPAMINE

The sequelæ of serotonergic inhibition on the average value and the average prediction error are interesting in the light of the complex relationship between dopamine and serotonin. Phasic dopamine is known to report a signal related to the prediction error (Montague et al., 1996; Schultz et al., 1997), while the tonic levels of dopamine may report something more akin to the average reward (Niv et al., 2005, 2006; Floresco et al., 2003; Goto and Grace, 2005). High levels of serotonin in this model effectively lead to higher levels of average expected reward, and thus this might predict an overall synergistic effect of endogenous dopamine and serotonin levels. A synergistic effect of serotonin and tonic DA levels is observed with microdialysis in the NAcc (Galloway et al., 1993; De Deurwaerdere P, 1998; Parsons and Jr, 1993). On the other hand, infusions of 5HT into limbic centres are known to inhibit ongoing (appetitive) behaviour (Kapur and Remington, 1996; Fletcher and Korth, 1999), which is in keeping with the basic tenet of this model, whereby 5HT inhibits actions. Reductions in 5HT lead to a reduction of average reward and may thus result in lower levels of tonic dopamine, although the present work makes no detailed prediction about phasic dopamine, and thus about data that show altered acquisition of reward-related responses.

One interesting alternative view of 5-HT due to (Doya, 2000) is that it is involved in controlling the appropriate timescale of behaviour by determining the discount factor for future affective outcomes (parameter γ in equation 5.1). In this theory, 5-HT depletion reduces the effective value of γ, making subjects appear more impulsive (Tanaka et al., 2004; Schweighofer et al., 2006, 2007). Our model captures impulsivity in a different way, by specifically facilitating the choice of aversive actions rather than changing the timescale of evaluation.

5.4.4 DEPRESSION

It first has to be pointed out that it is unclear whether the present model is more relevant to depression or to anxiety. Firstly, this is because there is no thorough definition of either disease in terms of reinforcement mechanisms. There is also at best a fuzzy distinction between the two in terms of risk factors (Hettema et al., 2006b) and pharmacology (Ressler and Nemeroff, 2000), and they are extraordinarily comorbid (Kaufman and Charney, 2000). We will conjecture a dichotomy in chapter 6.

While TrD is a very reliable way of re-inducing depression (but not anxiety), it is not the only one. For instance, patients who are responsive to SNRIs are more sensitive to α-methyl-tyrosine (see section 2.1.4) than TrD and a recent report with a DA antagonist successfully re-induced depressive symptoms in formerly depressed people (Willner et al., 2005). The latter authors suggest that DA may be a "final common path" for depression, and may relate more to the depressive state than serotonin, which in turn may be more important in defining a trait

(Willner, 1985b; Heinz, 1999; Willner, 2002).

We would like to follow this suggestion and conjecture the following: that the depressive state is characterised by low DA *as a result of* low 5HT. However, we see this not as a direct interaction between the neuromodulators, but as a signature of interaction between the Pavlovian, the goal-directed and the habitual affective decision making systems. Assume a drop of serotonin (without assuming what precipitates this drop), as described in this chapter, in an individual who usually relies on serotonergic inhibition. Assume further that the drop itself goes unnoticed, but that its consequences (unexpected punishments, large negative prediction errors, a drop in average reward) do not. This change in reward statistics needs explanation. We suggest that other affective systems ascribe it to a shift in the environment, and cause *normative* behavioural responses. The unexpected punishments might be interpreted as a lack of control (as described in chapter 4), with attendant dopaminergic repercussions. Alternatively, it may be that the goal-directed system uses a variable akin to the average expected reward for actions as an approximation to an estimate of control. A change in the average reward could then cause the changes in expected control. Successful accounting of the apparent alteration of reward statistics might then contribute towards stabilising the original changes in the serotonergic system.

The present model does not describe *why* there should be a drop in $\alpha_{5\mathrm{HT}}$. One option is a process at a purely biological level, such as invoked by TrD, or maybe some pathological process. While this chapter was formulated as if the change in 5HT were imposed in such a manner from without the system, we are certainly not wedded to it. Rather, there are various ways in which this may be achieved, with similar consequences. One alternative option is as a normative response to changes in environments' reward statistics, such as those experienced by animals in LH, CMS or other behavioural models of depression. However, for this one would need a more general theory of inhibition — what level of inhibition is optimal? Tools for the characterisation of the trade-off between accurate knowledge about a state's value and the cost incurred in learning about it are already in existence (Baum and Smith, 1997; Dearden et al., 1998, 1999) and might be applicable to aspects of the present scenario.

VI

General Conclusions

6.1 Contributions

The aim of the thesis was to illustrate that affective decision making can serve as an integrative framework for the various approaches taken to depression. By extrapolation, it is hoped that such a demonstration might indicate the usefulness of affective decision making to other psychiatric disorders. We argue that, while the biological, behavioural and cognitive data contribute differentially to our understanding of the involvement of different aspects of affective decision making, the fact that the major findings of these three divergent approaches can to a large extent be cast within that one structure in itself provides strong evidence for the usefulness of the framework. Five aspects of affective decision making guided the first part of our review on the characterisation of the depressed state, while the second part concentrated on the induction of depression in humans and animals. Briefly, the following is apparent in the literature:

1. Changes in primary reinforcer sensitivity: There is converging evidence that the state of depression brings a symmetric, blunted sensitivity to specific primary punishments and rewards. However, it brings a heightened sensitivity to stress, and stress itself influences the perception of rewards. In animal models, induction of a depressed state by exposure to uncontrollable stressors is usually followed by both analgesia and impairments of the dopaminergic reward system.

2. Goal-directed decisions: The depressed state is characterised by a perception of no control which implicates the goal-directed decision-making system. There is at present no unambiguous human behavioural data to support this. In animals, the strongest evidence comes from the behavioural generalisation of learned helplessness across reinforcer valence (see chapters 3 and 4) and the sensitivity to prefrontal lesions (Amat et al., 2005).

3. Pavlovian decisions: The involvement of serotonin in depression suggests that the state of depression is associated with decreased inhibition of actions that lead to negative outcomes.

4. Habitual decisions: There is no specific evidence that habitual decision making is altered in depression. However, there is behavioural data from habitual paradigms that can be either explained by a change in primary reinforcer sensitivity, or by alterations to habitual learning.

5. Motivation: There is evidence that a subgroup of patients have decreased general motivation (and tonic DA levels), but there is evidence that the motivational processes *per se* are unimpaired.

The main body of the thesis uses computational models to explore the points 1-3 above in detail. Given its central status in research on depression, we first approach learned helplessness (LH). Generalisation is at the heart of LH, but the question is generalisation of what? In chapter 3 we examine the importance of generalising analgesia, a blunting of primary reinforcer sensitivity for which there is also evidence in human depression. We found that a simple formulation in which blunting lessened the impact of shocks but incurred an evolutionarily fixed cost maintained the well-posed nature of the reinforcement problem and allowed extensive aspects of LH to be reproduced. Furthermore, it had important consequences for the continued acquisition of optimal actions. Thus, blunting may play a role in the maintenance of depression and it is important to keep effects potentially due to primary reinforcer changes in mind when analysing the other aspects of affective decisions.

Chapter 4 presented an explicit formulation of control. This naturally accounts for the generalisation of helplessness effects across reinforcer valence. Generalisation of the control variable itself on the other hand bring CMS within the scope of goal-directed action choice; gives insight into an important determinant of the inter-individual variability in the sensitivity to uncontrollable reinforcement; and as such provides a concise formalisation of the notion of internal/global/stable versus external/specific/unstable attributions. It is hoped that this will facilitate behavioural investigations of goal-directed decision making, particularly with respect to helplessness, in humans.

Finally, chapter 5 took three critical facts about serotonin and argued that they and their *prima facie* contradictory nature can be understood as related to a particular computational strategy (pruning) in reinforcement learning. The Pavlovian system is argued to be simple, have access to a fixed (though not entirely destitute) set of actions, and to be evolutionarily ancient and dominant. We showed how reliance on one class of Pavlovian actions — inhibition of actions with low expected future outcomes — while in general a desirable strategy, yields overly optimistic value functions. It produces a vulnerability to a sudden drop of inhibition, resulting in large unexpected punishments and a sharp drop in the average reward earned.

6.2 LIMITATIONS

There are other important aspects to depression, such as vegetative ones, which we so far have neglected. Briefly, our approach is to attempt to understand the interrelationship between and the neurobiology of the affective aspects first, and it is hoped that the link to neurobiology will then allow the incorporation of these other aspects. For example, it may be that a fuller understanding of the involvement of serotonin may explain why some patients suffer from hypersomnia and others from insomnia. However, this remains to be shown.

Other major limitations to this work remain. We have neglected vast areas of research. There is extensive work on many more animal models, both behavioural and biological. We have neglected noradrenaline entirely, although it may be as efficacious in the treatment of depression as serotonin. Only very general information was derived from treatment. There is by now extensive neuroimaging data on depression, which, together with neuroimaging data in normals, is very likely to bear on the issues treated here. Furthermore, it should be reiterated that our arguments, particularly the human behavioural ones, are based on sparse amounts of data that were not designed to test these hypotheses. We have vastly oversimplified even those aspects of depression that we did model. Finally, we have, as far as possible neglected all aspects of the data that are affected by issues of consciousness, and may thus have failed to provide an account that matches the prominent subjective features of depression. However, our main aim was a proof of principle, and so we concentrated on what we judged the most directly relevant data. It will be very important to attempt to incorporate these other types of data in the future.

6.3 FUTURE WORK

The work presented here is but the beginning of much more work that has to follow if these ideas are to carry fruit. We pointed out before that some of this work has been undertaken to facilitate the development of behavioural tasks to assess the affective system in depression. In humans, we have begun to design simple choice tasks and assessments of primary reinforcer value. Depressed subjects will undergo a battery of tests specifically designed to isolate different types of affective decisions. First, this will give direct information on the separate functioning of the systems, and for example answer whether goal-directed learning is, as introspection suggests, really affected. Secondly however, it is hoped that it will be informative about the joint functioning of the various systems in individuals and provide evidence for or against the kind of interaction we speculated might induce and / or maintain depression. Furthermore, the subject group will include cases of pure depression but also depression with co-morbid psychiatric disorders, and this might give insights into the functional issues that underlie the extensive co-morbidity, prominently with anxiety.

In addition to these experiments that test the contents of generalisation, it is important to test generalisation itself in depression. Such experiments are straightforwardly adapted from e.g. the literature on generalisation of fear conditioning. In animals, such experiments may usefully be combined with manipulations of the hippocampus given both the recent evidence

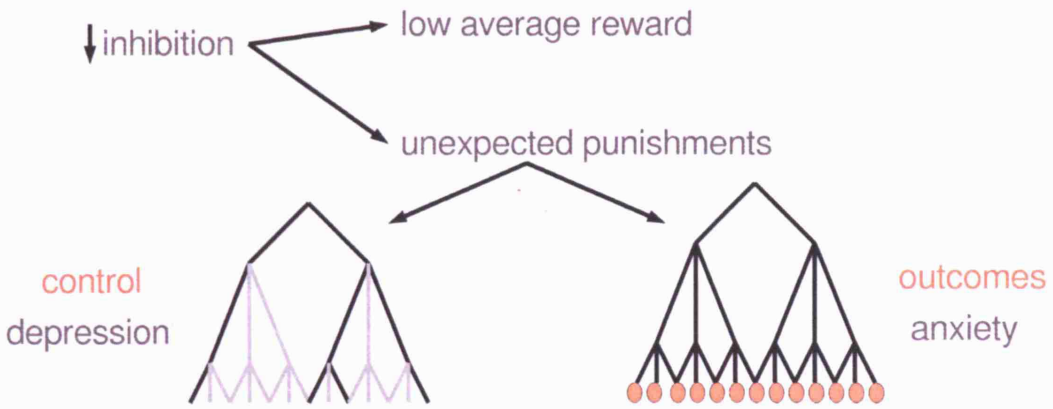


FIGURE 6.1: Serotonergic basis of anxiety and depression comorbidity. A decrease in Pavlovian inhibition might be interpreted by the goal-directed system in a variety of ways and give rise to different psychiatric disorders. For example, the sudden increase in unexpected punishments can be interpreted in at least two ways. It can be seen as evidence for a lack of control, which as we saw can be expressed in terms of priors on the decision tree (left). Alternatively the changes in reinforcements obtained could be due to a downwards shift in the reinforcements available in the environment (right). The former is suggested to lead to depression, the latter to anxiety. This may be one reason for the extensive comorbidity between anxiety and depression.

on the importance of hippocampal neurogenesis to the function of antidepressants (Santarelli et al., 2003) and also data on the importance of e.g. hippocampal $5HT_{1A}$ receptors (Graeff et al., 1998).

We would like to close with one comment on the relationship between psychiatry and normativity explored here. It is clear that psychiatric disorders are maladaptive, and we are not arguing that it serves a particular purpose *per se* (Nesse, 2000; Stevens and Price, 2000). However, the high prevalence of psychiatric disorders, especially of depression, has to us been a strong indicator that the maladaptivity arises from fundamental constraints on the brain and the tasks it faces, i.e. it is the signature of a trade-off that allows the brain to function adaptively in most circumstances. It is this strong determination by normal function which we have started to explore here, and which we believe carries great promise as it may, just may, provide the aetiological link between psychiatric dysfunction and normal function.

6.4 SYNTHESIS

To conclude, we return to our the introduction, where we saw that the different affective decision-making systems rely on different computational solutions to the reinforcement problem. These systems are used in parallel by animals (Killcross and Coutureau, 2003), and animals take the advantages and disadvantages (Sutton and Barto, 1998) of each system into account when they arbitrate between them (Daw et al., 2005). However, it is also clear that non-optimal behaviour can at times result from the interaction between the systems (Dayan

and Balleine, 2002; Dayan et al., 2006). Based on this, we speculate that a predisposition to depression arises from a constitutively overactive Pavlovian inhibitory system, leading to an optimistic appreciation of the world. An acute episode of depression is hypothesised to arise from a sudden drop in the Pavlovian system's inhibition of actions with negative expected outcomes. This sudden drop leads to large, frequent and unexpted punishments, negative prediction errors and a drop in the average reinforcement rate. Stabilisation of the depressive state can then result from a misinterpretation of a change in the functioning of the Pavlovian system as an environmental reinforcement shift. I.e., rather than attributing the sudden deluge of unexpected punishments to a deficiency of its own Pavlovian system, the brain assigns it to a change in the structure of the reinforcement problem it faces. This change in the reinforcement structure is now learned by fully functional, healthy habitual or by the goal-directed systems. This is because the Pavlovian system is hypothesised to learn on a much longer (evolutionary) timescale, whereas the habitual, and yet more so the goal-directed system, learn on the timescale on which environmental contingencies do change. Once the change, which is due to an internal malfunction, is accounted for in this manner, behaviour can again stabilise, as might the depressive state.

In particular, we suggest that the unexpected punishments experienced after a sudden drop of 5HT might be interpreted as evidence for no control by the goal-directed system (because punishments suddenly appear not to be avoidable any more); as evidence for an environmental drop in the average reward rate, resulting in a lowered motivation. Furthermore, if an organism is (apparently) punished for all behaviours, then punishments become less informative for optimal action choice, and blunting may be appropriate. However, we have also been at pains pointing out that the data on serotonin does not exclusively associate it with depression. As such, one may further speculate that a 5HT drop might also be interpreted by some other system as evidence that the actual reinforcers available in the environment have become more negative. Such a prediction that outcomes, even of controlled actions, are likely to be negative, might be more associated with anxiety. Let us reiterate: a downwards shift in the experienced reinforcers, due to a change in one of the systems subserving action choice, or due to a true environmental change, or both, might be interpreted either as a change in the tree structure (control, depression), or in the value of the leaves of the tree (negative outcome predictions, anxiety), or as a mixture of the two (figure 6.1). While this is extremely speculative, we hope that it gives some intuition to the kinds of interactions between aspects of the affective system we envisage might explain the comorbidity between different diseases, and thus be helpful in their classification.

APPENDIX

A

REINFORCEMENT LEARNING AND AFFECTIVE DECISIONS

To begin with, we need to specify precisely the structure of affective decision making we will be concerned with, i.e. we need to set up the framework of normative affective decision making and its neurobiological basis. The credo goes as follows: affective decisions are guided by the attempt to maximise positive affect. As we strongly believe that consciousness is still far beyond the reach of science, we will concentrate on a very specific interpretation of affect stripped bare of all conscious, subjective colourings. When talking about hedonia, really we will exclusively focus on reinforcement (Berridge and Robinson, 1998). Where this is not possible, we will attempt to push as far as possible in this direction. Thus, we are concerned with decisions that lead to maximal reinforcement. This is precisely the domain of reinforcement learning: Given choices made in the past led to a set of reinforcements, what future choices should be made to maximise the total future outcomes?

Consider first an overly simple scenario, where choice of action a leads to outcome A on a fraction f_a of the times a is chosen, and choice of action b leads to outcome B on a fraction f_b of the time b is chosen. The rest of the time, no outcomes are observed for either choice. The better of the two choices is that with the greater expected outcome. The expected outcomes are $\mathbb{E}_a = Af_a$ and $\mathbb{E}_b = Bf_b$. The choices thus depend both on the probability with which the outcome is estimated to occur, and on how desirable the outcomes are. Contemporary theories of depression implicitly argue about either one or the other of these two factors: either there is some change in the desirability of outcomes (theories that state "depression=anhedonia"), or the probability that they will be observed is judged differently (theories that state "depression=learned helplessness").

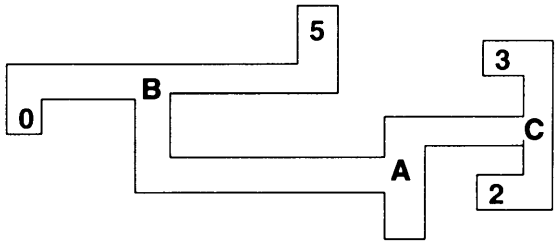


FIGURE A.1: A prototypical reinforcement learning setting. States are positions within the maze. Start at state A, and choose whether to turn right or left. This leads either to state B or state C, but no rewards as yet. At state B and C another choice has to be made, leading finally to the states at the end of the maze, which are armed with rewards in red. The optimal policy is to first turn left, then turn right. Adapted from Dayan and Abbott (2001).

A.1 REINFORCEMENT LEARNING

We here give a very concise overview of the fundamental reinforcement learning techniques used in the thesis. Let us define the reinforcement learning framework in greater generality. A Markov decision process (MDP) consists of the following:

- States $s \in \mathcal{S}$
- Actions $a \in \mathcal{A}$
- For each action a, a transition matrix $\mathcal{T}^a$. The entry $\mathcal{T}^a_{ij}$ gives the probability $p(s_j|s_i, a)$ of moving from state i to state j when taking action a.
- For each state-action-state triple, there is a reward. It's expectation is $\mathcal{R}(s, a, s')$.
- The solution of a MDP is a policy $\pi(a|s)$. A policy is a distribution over actions for each state.

In words, one starts at some state s, chooses some actions a, which leads to a new state s' according to the transition matrix $\mathcal{T}^a$, and which yields a reward according to the reward structure $\mathcal{R}$ (which may be probabilistic). Figure A.1 gives a very simple example in which there are seven states, two actions with deterministic transitions, and the rewards depend only on the state $\mathcal{R}(s, a, s') = \mathcal{R}(s)$. A policy π assigns an action to each state. Obviously, it makes sense to go left at A, right at B and left at C. Reinforcement learning is concerned with how such a policy can be inferred from sparse information gleaned about the environment while exploring it. That is, the subject makes choices and observes the outcomes (the next states, and the reinforcements). The policy sought is that which maximises the total expected reward. Consider a slight variation of the figure A.1, in which state B yields reinforcement -1. The best policy is still to go left and then right. But if we were to maximise the reward obtained at each state separately, we would choose to go right at state A, resulting in a suboptimal overall policy. The reinforcement learning problem is mainly difficult because it attempts to maximise the *total* reward, rather than the individual local rewards. Additionally, if the subject had always chosen

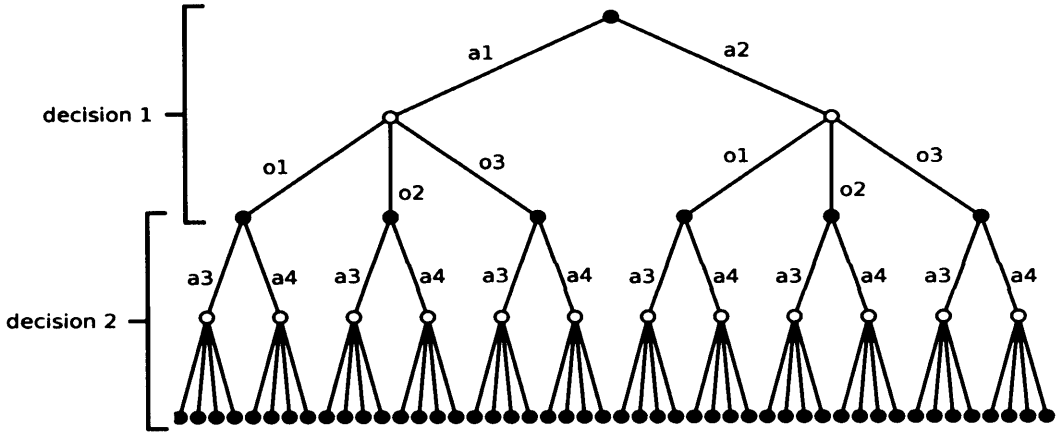


FIGURE A.2: A full decision tree for a Markov decision process of depth $D = 2$, with 2 actions at each stage (the first decision is a choice between $a1$ and $a2$, and the second between $a3$ and $a4$) . After the first decision, there are three possible outcomes ($\{o1, o2, o3\}$), after the second one four. The semantics are as follows: After each observation (black nodes), one of two actions (white nodes) has to be chosen. Each of the actions in turn leads to one outcome (black or grey observations nodes), with associated (known) rewards. The final observation nodes are grey. A path ξ through the tree is a set of actions and outcomes from the root to one of the grey leaf nodes $\xi_d = (a_d, o_d, \cdots a_D, o_D)$. The optimal first action at the root of the tree is that which maximises the reward obtained along the entire path, rather than that which maximises the immediate reward from the next outcome. An explicit solution of the MDP problem evaluates all paths and chooses the one with the highest expected outcome.

to go left at B, he would not know that there is a luscious 5 units of reinforcements lurking in the right arm, and so the best policy based on what he knows is not actually the best overall policy: there is always a trade-off between exploring, in case there are large, as yet unobserved, rewards, and exploiting what one knows.

The goal of reinforcement learning is to infer the best policy. The methods used here all work via a value function: some policy is fixed, and under that policy, each state's total future expected reward — the value of that state — is computed. Based on this information, the policy is updated. The methods differ in the amount of prior knowledge assumed about the MDP and in whether they do a full policy evaluation, or just a partial one.

A.2 VALUE FROM TREE SEARCH

The straightforward solution depends on knowing both what states should be visited ($\mathcal{R}$), and what consequences actions have (this is the "model of the world", $\mathcal{T}$). All actions and their their consequences can then be simulated to evaluate their value. More specifically, let us assume we

follow some policy π for D choices. In combination with the transition matrix $\mathcal{T}$, this defines a probability distribution $p(\xi)$ over paths $\xi \equiv (s_0, a_0, s_1, a_1 \cdots s_D)$ given by

$$p^{\pi}(\xi|\mathcal{T}) = p(s_0) \prod_{d=0}^{D} \pi(a_d|s_d) \prod_{d=0}^{D-1} \mathcal{T}(s_{d+1}|s_d, a_d) \tag{A.1}$$

Figure A.2 shows all possible paths for a small MDP in which $D = 2$, $|A| = 2$ and each action leads to one of three or four different outcomes / states. The total expected reward starting at some state $s = s_0$ is then simply

$$\begin{aligned} V^{\pi}(s|\mathcal{T}) &= \sum_{\xi} p^{\pi}(\xi|\mathcal{T})\mathcal{R}(\xi) \\ \text{where} \quad \mathcal{R}(\xi) &= \sum_{d=0}^{D-1} \mathcal{R}(s_d, a_d, s_{d+1}) \end{aligned} \tag{A.2}$$

To evaluate this, one sums the rewards along all possible paths, and then averages over the probability of that path — for every decision. Thus, optimal choice of policies (action sequences) still has two components: the likelihood of outcomes, and their size. If it is feasible (in very small problems), this is the exact, model-based solution to a MDP. We will review some evidence that indicates that this strategy underlies goal-directed behaviour (Daw et al., 2006).

Usually, there is no or only incomplete knowledge about the effects of actions (i.e. $\mathcal{T}$ is unknown). This is the situation we are interested in with respect to theories of depression. Consider the scenario in which $\mathcal{R}$ is known and $\mathcal{T}$ is unknown, but transitions between states are observed and collected in matrices $\mathbf{N}$. If we have access to some model (such as a Dirichlet prior with parameters θ) that relates observations to underlying transition matrices, we can use the model and the observations to furnish a distribution over transition matrices $p(\mathcal{T}|\mathbf{N}, \theta)$, and recover an analogue of equation A.2. Chapter 4 will be concerned mainly with these kinds of models. We will argue that the formalism captures the major aspects of the concept of control in the learned helplessness and associated literatures, and that it allows us to replicate the main findings. This formalism links control tightly to goal-directed behaviour and as such makes clear, testable predictions. It also has implications for motivational theories of depression. Briefly, high control will be translated into a setting of θ such that $\pi(s|a)$ tends to have low entropy. When we talk about control in chapter 2, we will have this in mind, although a precise definition will have to wait until chapter 4.

A.3 MODEL-FREE ESTIMATES OF VALUE

The number of paths grows with $D^{|\mathcal{A}||\mathcal{S}|}$, where $|\mathcal{A}|$ is the number of actions at each state. Due to the decomposition in equation A.1 (the Markov property), the sum can be expanded and

equation A.2 reformulated

$$
\begin{aligned}
V^{\pi}(s_0) &= \sum_{a_0} \pi(a_0|s_0) \sum_{s_1} p(s_1|s_0,a) \Big[\mathcal{R}(s_0,a_0,s_1) + \dots \\
&\quad \underbrace{\sum_{a_1} \pi(a_1|s_1) \sum_{s_2} p(s_2|s_1,a) \left[\mathcal{R}(s_1,a_1,s_2) + \cdots \right]}_{=V^{\pi}(s_1)} \Big] \\
\Rightarrow V^{\pi}(s) &= \sum_{a} \pi(a|s) \sum_{s'} p(s'|s,a) \left[\mathcal{R}(s,a,s') + \gamma V(s') \right] \qquad \text{(A.3)}
\end{aligned}
$$

where a discount factor $0 < \gamma < 1$ has been added which weighs rewards in the distant future less than proximal ones. Equation A.3 is only exact for $D \to \infty$ and is known as the Bellman equation (Sutton and Barto, 1998). It is a fixed-point equation (the same quantity appears on the left and the right hand side of the equation) and can be solved by iteratively updating one state after the other until convergence.

The Bellman equation (A.3) assumes knowledge of both $\mathcal{T}$ and $\mathcal{R}$ and uses it to explicitly perform averages over the outcomes of all actions. Cached methods such as TD learning and the simpler Δ rule (Shanks, 1995; Dayan and Abbott, 2001) are approximations to the Bellman equation that dispense with explicit models. Assume that only sample paths are given, and replace the averages and the reward by a single sample (s_d, a_d, s_{d+1}) and $r_d = r(s_d, a_d, s_{d+1})$. The Bellman equation states that, on average, $V(s_d) = r_d + \gamma V(s_{d+1})$, i.e. that the value of a state is, on average, the sum of the immediate reward r_d and the value of the next state. Again, this is a fixed-point equation, and we can write update equations based on the discrepancy δV

$$
\begin{aligned}
\delta V &= \gamma V(s_{d+1}) + r_d - V(s_d) \qquad \text{(A.4)} \\
V(s_d) &\leftarrow V(s_d) + \varepsilon \delta V \qquad \text{(A.5)}
\end{aligned}
$$

If the number of samples equals $1/\varepsilon$, and the policy π is kept constant, this is a sampling approximation to the average in equation A.3. It is worth commenting on this equation. δV is the so-called prediction error. It is the difference between the expected reward $(\gamma V(s') - V(s))$ and the obtained reward. TD is essentially a simple Kalman filter. Rather than storing all past rewards and performing an average every time the information is required, this updates the parameter of interest (the predicted total reward) on-line. We will see that cached representations of reward are closely associated with habits.

A.4 POLICIES FROM VALUES

If the above policy evaluation steps are run to convergence, $V^{\pi}(s)$ is known. A new policy is then obtained by choosing, for each state, the action which leads to the next state with the largest value:

$$
\pi(s) \leftarrow \arg\max_{a} \sum_{s'} p(s'|a) V(s') \qquad \text{(A.6)}
$$

The policy iteration algorithm alternates between the policy evaluation in equation A.3 and policy update. It is proven to converge to a global optimum (Bertsekas and Tsitsiklis, 1996).

Alternatively, one can update the policy before the full sweep, or make it a function of the values themselves. This is used in Q learning, and in particular for SARSA. SARSA is an algorithm with convergence guarantees which is named after its use of samples from $(s_d, a_d, r_d, s_{d+1}, a_{d+1})$. Father than constructing just $V(s)$, it constructs $Q(s,a)$, the value of taking action a in state s. For a given policy, $V(s) = \sum_a \pi(a|s)Q(s,a)$. The TD equations can equally be written for Q values (Watkins and Dayan, 1992):

$$\delta Q = \gamma Q(s_{d+1}, a_{d+1}) + r_d - Q(s_d, a_d) \quad \text{(A.7)}$$

$$Q(s_d, a_d) \leftarrow Q(s_d, ad) + \varepsilon \delta Q \quad \text{(A.8)}$$

and let the policy be a softmax-ed version of the Q values:

$$\pi(s|a) = \frac{\exp(\beta Q(s,a))}{\sum_{a'} \exp(\beta Q(s,a'))} \quad \text{(A.9)}$$

where β sets how strongly behaviour is dictated by the Q values. When $\beta = 0$, the values have no effect on behaviour, whereas when $\beta \to \infty$ the action with the strictly maximal value is deterministically chosen. Thus, it may be possible to learn values normally, but simply not act on them.

In the Bellman equation, we sneakily introduced $0 < \gamma < 1$. Without this factor, the total future reward for an infinitely long action sequence might well be infinite. Average reinforcement learning (Mahadevan, 1996) instead focuses on the advantage of one action over others. The dQ value of an action is defined as

$$\delta dQ = dQ(s_{d+1}, a_{d+1}) + r_d - dQ(s_d, a_d) - \rho(s_d) \quad \text{(A.10)}$$

$$\rho(s_d) = \rho(s_d) + \varepsilon(r_d - \rho(s_d)) \quad \text{(A.11)}$$

i.e. ρ is the running average of the reward received when visiting state s and the advantages $dQ(s,a)$ of actions in that state are defined with reference to it. Over time, as the best action comes to always be chosen, the advantage of the optimal action decreases to zero, while all other actions end up with negative advantages. If the measure of average reward ρ were under independent control, it would change the advantage of actions.

A.5 Decision tree

The exact solution of a MDP is achieved by iterating through a decision tree (see figure A.2 for an example). Assume d actions have been taken and there are $D - d$ actions left, i.e. we are at depth d in a tree of total depth D. The observations $\mathbf{N}_d$ up to depth d together with the models described in section B allow us to derive, recursively, for each action a, a probability distribution over the immediate outcomes $p(o^a_{d+1}|\mathbf{N}_d, \theta)$ where θ are the control parameters of the particular model. The immediate expected outcome for taking action a is then simply $r(d,a) = \sum_j R_j p(o^a_{d+1} = j|\mathbf{N}_d, \theta)$. We would like to choose the best action in terms of long-

term outcomes, i.e. the action with the highest expected reward over the entire remaining tree. This can be expressed iteratively. The expected total reward $Q(a)$ for an action is its immediate reward, plus the reward from subsequent optimal action choices:

$$Q(a, d) = r(d, a) + \max_{a'} Q(a', d + 1) \tag{A.12}$$

Thus, the models specified in section B can be seen as distributions over such trees. High control corresponds to the assumption that the trees branch little relative to low control. Even so, the trees rapidly becomes too large to compute — the number of paths through the tree grows as D^{AL}. We thus restrict ourselves to small trees in this section. POMDPs

B

STATISTICAL DESCRIPTIONS OF CONTROL

We here give the mathematical details of the various models of control described in chapter 4: outcome entropy; fraction of controllably achievable outcomes and fraction of controllably achievable reinforcement.

Briefly, the general setup is the following: Environments are assumed to be characterised by particular levels of control, i.e. the likelihood of observations is parametrised according to some suitably defined control parameter. Organisms collect observations in one (or a few) training environments, and based on this infer a posterior distribution over the setting of the control parameter in the training environments. Organisms are then transferred to a test environment and exposed to a limited number of observations. Organisms combine their prior expectations about the level of control in the test environment (derived in an again suitable manner from the posterior distributions over control in the training environments) with the likelihood of the observations in the test environment and arrive at a predictive distribution for future observations in the test environment. Actions in the test environment are chosen according to the predictive probabilities of outcomes.

B.1 CONTROL AS CONDITIONAL ENTROPY / OUTCOME SET SIZE

The first and most basic notion of control is that of the entropy of the probability distribution over outcomes, conditioned on an individual action (Maier and Seligman, 1976; Overmier et al., 1980; Gibbon et al., 1974).

Let us first just investigate the effect of outcome set sizes of independent actions, i.e. the number of outcomes that are potentially observable for any one action. The outcome set size is

related (though not equal) to the conditional entropy, but is analytically much more convenient. We follow the work of Friedman and Singer (1999); Dearden et al. (1998, 1999) closely. The setup is thus the following: given a number of action-outcome observations, and a prior belief about how many *different* observations are likely to be observed, what is the optimal action choice? The optimal action choice will be derived from the predictions about which outcomes are likely for the action.

Let us first consider a single action, with L possible outcomes. Let X be an unordered subset of these outcomes and $|X|$ be the cardinality of that set, i.e. the number of different elements in the set, e.g. for the subset $X = \{1, 2, L\}$ (for $L > 2$), $|X| = 3$. There are $\binom{L}{|X|} = L!/(|X|!(L - |X|)!)$ such sets of a given size for a total number of L outcomes. We will now put a prior distribution $p(|X|)$ on the size of the outcome set, i.e. on the number of different outcomes expected for a particular action, and assume that all sets of the same cardinality have equal probability. This leads to a prior on sets

$$p(X) = \binom{L}{|X|}^{-1} p(|X|) \tag{B.1}$$

Let us furthermore parametrise the prior on set size in equation B.1 as a truncated geometric distribution with parameter ζ:

$$\begin{aligned} p(|X||\zeta) &= \begin{cases} 1/L & \text{if} \quad \zeta = 1 \\ \zeta^{|X|-1} \frac{1-\zeta}{1-\zeta^L} & \text{else} \end{cases} \\ p(\zeta) &= \text{Gamma}(\alpha_\zeta, \beta_\zeta) \quad \text{s.t} \quad p(\zeta = 0) \approx 0 \\ p(X) &= \binom{L}{|X|}^{-1} \int_0^1 d\zeta \, p(|X| \, |\zeta) p(\zeta) = \binom{L}{|X|}^{-1} p(|X|) \end{aligned} \tag{B.2}$$

where as $\zeta \to -\infty$ only set size 1 is allowed, and as $\zeta \to \infty$ all but set size L is prohibited. Thus, the parameter ζ determines the set size, and is our parametrisation of control for this section.

To illustrate the pure effect of a prior on outcome size, we need to integrate out the effect of the actual probability distribution over that set. Let $\mathbf{c}$ denote the outcome probability vector of an action, i.e. the probability of observing outcome i is c_i, and the likelihood of observing outcome i n_i times is a multinomial

$$p(\mathbf{n}|\mathbf{c}) = \frac{(\sum_i n_i)!}{\prod_i n_i!} \prod_i c_i^{n_i} \tag{B.3}$$

It is now possible to put a Dirichlet prior, parametrised by the outcome set size $|X|$, on the multinomial vector of outcome probabilities $\mathbf{c}$:

$$p(\mathbf{c}|X, \alpha) = \frac{\Gamma(|X|\alpha)}{\prod_{i \in X} \Gamma(\alpha)} \prod_{i \in X} c_i^{\alpha - 1} \tag{B.4}$$

$$\tag{B.5}$$

which put mass on vectors $\mathbf{c}$ with $|X|$ nonzero elements. We let α be relatively large to ensure that all outcomes in X have a large probability of actually generating data (putting most probability mass on vectors $\mathbf{c}$ such that $c_i \approx c_j \, \forall i, j \in X$). The predictive probability that the

outcome at the next action $D+1$, given that D outcomes have already been observed, is a standard multinomial as a Dirichlet prior is conjugate to the multinomial:

$$p(n_{D+1} = j|\mathbf{n}, X, \alpha) = \begin{cases} \frac{\alpha+n_j}{|X|\alpha+N} & \text{if } j \in X \\ 0 & \text{else} \end{cases} \tag{B.6}$$

Note importantly, that this only applies to outcomes *within* the set X on which we condition. Given our prior over sets in equation B.1, this allows us to derive the probability of observing any outcome by averaging over set sizes. Note however, that sets that do not contain the set of previously observed outcomes (call this set Y) have zero likelihood and thus do not contribute to the predictive distribution:

$$p(n_{D+1} = j|\mathbf{n}, \alpha) = \sum_{X \supseteq \{Y,j\}} p(n_{D+1}|\mathbf{n}, X)p(X|\mathbf{n}, \alpha) \tag{B.7}$$

$$\begin{aligned} p(\mathbf{n}|X, \alpha) &= \int d\mathbf{c} p(\mathbf{n}|\mathbf{c})p(\mathbf{c}|X, \alpha) \\ &= \frac{N!}{\prod_{i \in X} n_i!} \frac{\Gamma(|X|\alpha)}{\Gamma(|X|\alpha + N)} \prod_{i \in X} \frac{\Gamma(\alpha + n_i)}{\Gamma(\alpha)} \end{aligned} \tag{B.8}$$

$$p(X|\mathbf{n}, \alpha) = \frac{p(\mathbf{n}|X, \alpha)p(X)}{\sum_X p(\mathbf{n}|X, \alpha)p(X)} = \frac{B(X)}{\sum_{X \supseteq Y} B(X)} \tag{B.9}$$

$$B(X) = \frac{\Gamma(|X|\alpha)}{\Gamma(|X|\alpha + N)} \prod_{i \in X} \frac{\Gamma(\alpha + n_i)}{\Gamma(\alpha)} \binom{L}{|X|}^{-1} p(|X|)$$

$$\Rightarrow p(n_{D+1} = j|\mathbf{n}, \alpha) = \frac{\sum_{X \supseteq \{Y,j\}} \frac{\alpha+n_j}{|X|\alpha+N} B(X)}{\sum_{X \supseteq Y} B(X)} \tag{B.10}$$

Equation B.8 is a standard Dirichlet integral, equation B.9 is the standard Bayes theorem and equation B.10 is the predictive distribution given D outcomes. As we will here mainly be dealing with problems in which L is small, say around 6, we can evaluate these sums explicitly, although it is straightforward to sample from the sets X that have nonzero likelihood.

For generalisation, given a set of observations, on several actions that share the setting of ζ, we will want to infer the maximum a posteriori value for ζ, which we can do via EM. Assuming

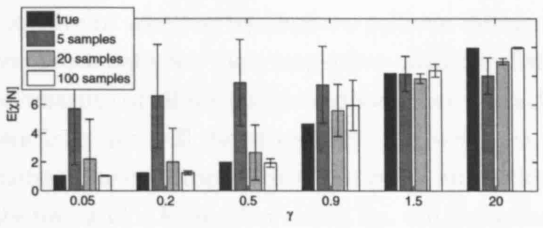


FIGURE B.1: Inferring ζ from observations on $L = 20$ independent actions with $L = 10$ possible outcomes each, averaging over c with $\alpha = 20$.

a high, fixed value for α,

$$\begin{aligned}
\hat{\zeta} &\equiv \arg\max_{\zeta} \log p(\zeta|\mathbf{N}) = \arg\max_{\zeta} [\log p(\mathbf{N}|\zeta) + \log p(\zeta)] \\
p(\mathbf{N}|\zeta) &= \prod_a p(\mathbf{n}^a|\zeta) = \prod_a \sum_{X^a} p(\mathbf{n}^a|X^a) p(X^a|\zeta) \\
\log p(\mathbf{N}|\zeta) &= \sum_a \log \sum_{X^a} p(\mathbf{n}^a|X^a) p(X^a|\zeta) \\
\Rightarrow \text{M step:} \quad 0 &= \frac{\partial}{\partial \zeta} \sum_a \langle \log p(\mathbf{n}^a, X^a|\zeta) \rangle_{q_a} + \frac{\partial \log p(\zeta)}{\partial \zeta} \\
&= \frac{\partial}{\partial \zeta} \sum_a \langle \log p(X^a|\zeta) \rangle_{q_a} + \frac{\partial \log p(\zeta)}{\partial \zeta} \\
&= \frac{1}{\zeta} \left(\sum_a \langle |X^a| \rangle_{q_a} \right) + L \left(\frac{1}{\zeta - 1} + \frac{L\zeta^{L-1}}{\zeta^L - 1} - \frac{1}{\zeta} \right) \\
&\quad + \frac{\partial \log p(\zeta)}{\partial \zeta} \qquad \text{(B.11)} \\
\Rightarrow \text{E step:} \quad q_a &\leftarrow p(|X|^a|\mathbf{n}, \zeta) \qquad \text{(B.12)}
\end{aligned}$$

Figure B.1 shows inference of ζ according to equation B.12. For large ζ, accurate inference is possible even when very few samples have been observed, but at low ζ the inference is much noisier. At low sample numbers, the likelihood appears to contain two modes, one at low, and one at high ζ, to account for the few cases in which 2 or more outcomes are observed for a particular action. The second mode however disappears rapidly with added sampling, or is eliminated by adding in even a weak prior (data not shown).

B.2 MULTIPLE ACTIONS WITH INDEPENDENT OUTCOMES

However, when multiple actions are considered, it is not sufficient to formulate control as relating to the entropy of the individual actions alone. There also needs to be some measure of the relationship between actions, and some notion of how many outcomes are favoured by a particular action. We now proceed to a definition of control that takes into account whether different actions achieve different outcomes, and whether these cover the range of outcomes possible in an environment. We then look at the consequences for the expected values of outcomes and for the exploration behaviour which are then investigated in a Bayesian framework.

The most intuitive measure of whether two actions achieve different outcomes is given by the Kullback-Liebler divergence between their respective outcome distributions. For more actions, it is given by some measure of all the pairwise divergences, which is a complex function. For mathematical convencience we will therefore only deal with a simplified set of outcome distributions. We parametrise the conditional distribution of one action very simply as a mixture of a uniform distribution and a Kronecker delta, i.e. we write the probability of a set of observations $\mathbf{n}$, n_i being the number of times outcome i has been observed following the choice of action a

$$P(\mathbf{n}|c,\mathbf{m}) = \frac{(\sum_i n_i)!}{\prod_i n_i!} c^{\mathbf{n}^\intercal \mathbf{m}} \bar{c}^{\mathbf{n}^\intercal (1-\mathbf{m})} \qquad \bar{c} = \left(\frac{1-c}{L-1}\right) \tag{B.13}$$

where $\mathbf{m} = [0 \ldots 0\,1\,0 \ldots 0]^\intercal$ is a vector of length L that henceforth designates *at most* one of the outcomes as the "controllably attainable" one for that particular action. The *scalar* variable c (not to be confused with the outcome probability vector $\mathbf{c}$ in the previous section) determines the mixing distributions. We will say that it regulates the degree to which the outcome is "controllably achievable". The outcomes not designated by $\mathbf{m}$ all have equal probability. $\mathbf{n}$ is the vector of outcome counts. L is the number of potential outcomes, and for simplicity we assume that the number of available actions is equally L (though it is straightforward to relax this).

For $c \to 1$, only one outcome (the one for which $m_i = 1$ is true) is observed, whereas as $c \to 1/L$, any outcome might be observed. The outcome entropy for that action

$$\mathcal{H} = -\sum_i p_i \log p_i = -c\log(c) - (1-c)\log\frac{1-c}{L-1} \tag{B.14}$$

is a strictly monotonically decreasing function of c for $L > 2$.

For a set of independent actions, we can write the likelihood of observations (assuming independent observations for different actions):

$$P(\mathbf{N}|c,\mathbf{M}) = \prod_a \frac{(\sum_i n_i^a)!}{\prod_i n_i^a!} c^{(\mathbf{n}^a)^\intercal \mathbf{m}^a} \bar{c}^{(\mathbf{n}^a)^\intercal (1-\mathbf{m}^a)} \propto \prod_{ij} C_{ij}^{N_{ij}} \tag{B.15}$$

where we have assigned the a^{th} column vector $\mathbf{m}^a$ of the matrix $\mathbf{M}$ to action a and the matrix C is defined below. $\mathbf{N}$ is a matrix consisting of the column vector observations for each of the actions. Let us clarify the meaning of $\mathbf{M}$ one more time: each column stands for one action, each row for one outcome. A unity entry in a column designates that outcome as the main outcome for that action. A goal-directed actor would chose that action in order to maximise the chances of obtaining that outcome. The variable c determines the probability of actually observing the designated outcome as opposed to any other one.

The second notion of control now becomes apparent, in the relationship between the columns of $\mathbf{M}$, i.e. between the controllably achievable outcomes of different actions. Consider the matrices $\mathbf{M}$ and their associated matrices $\mathbf{C}$, whose entry denotes the probability of outcome i

given action j was chosen $C_{ij} = p(outcome = i|action = j)$

$$
\begin{aligned}
\mathbf{M}_0 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \mathbf{C}_0 &= \begin{bmatrix} \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} \\ c & c & c & c \\ \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} \end{bmatrix} \\
\mathbf{M}_1 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \mathbf{C}_1 &= \begin{bmatrix} c & \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & c & \frac{1-c}{L-1} & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & \frac{1-c}{L-1} & c & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & \frac{1-c}{L-1} & \frac{1-c}{L-1} & c \end{bmatrix} \\
\mathbf{M}_2 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \mathbf{C}_2 &= \begin{bmatrix} c & \frac{1-c}{L-1} & 1/L & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & c & 1/L & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & \frac{1-c}{L-1} & 1/L & \frac{1-c}{L-1} \\ \frac{1-c}{L-1} & \frac{1-c}{L-1} & 1/L & c \end{bmatrix} \\
\mathbf{M}_3 &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \mathbf{C}_3 &= \begin{bmatrix} c/2 & c/2 & 1/L & \frac{1-c}{L-1} \\ \frac{1-c}{L-2} & c/2 & 1/L & \frac{1-c}{L-1} \\ c/2 & \frac{1-c}{L-2} & 1/L & \frac{1-c}{L-1} \\ \frac{1-c}{L-2} & \frac{1-c}{L-2} & 1/L & c \end{bmatrix}
\end{aligned}
\tag{B.16}
$$

These have very different implications. For large c, these matrices now exemplify various dimensions along which a putative control variable may change.

- $\mathbf{M}_0$: outcome 2 is attainable, but it is also the only one attainable. For $c \leftarrow 1$, all actions deterministically lead to outcome 2.
- $\mathbf{M}_1$: one action available for each of the outcomes. As $c \leftarrow 1$, all actions can deterministically attain their outcomes. In this case, all outcomes would be controllably achievable.
- $\mathbf{M}_2$: actions available for a fraction (here 3/4) of the outcomes.
- $\mathbf{M}_3$: actions lead to more than a unique outcome, even for high c this does not lead to full control. We will not consider this setting any further.

as $c \rightarrow 1/L$, the observations these matrices generate the same, flat, uncontrollable outcomes. Later, we will also consider the notion that control is "about" some particularly reinforcing outcome.

B.2.1 Control as fraction of controllably attainable outcomes

When more than a single action is considered, we thus need to take the relationship between actions into consideration as illustrated in equation B.16. We return to the simple case of equation B.15, constraining the matrix M to have one unit entry in each column and row. If there are L actions and L outcomes, there are $L!$ such matrices. For small L, the relevant integrals can be evaluated explicitly. We write the likelihood of observations as in equation B.15, and add a

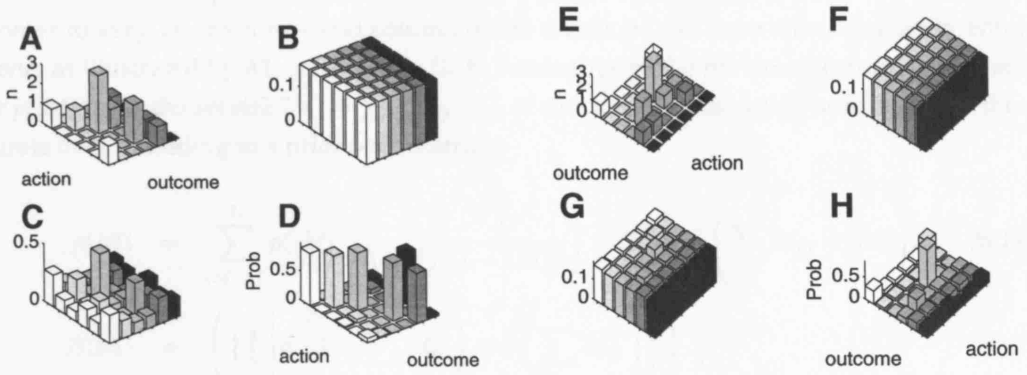


FIGURE B.2: Mean of posterior distribution over matrices $\mathbf{M}$ given data $\mathbf{N}$ in panel **A**. Panel **B** shows the posterior mean for $c = 1/L = 0.16$, i.e. no control at all, **C** for $c = 0.25$ and **D** for $c = 0.9$. Clearly, the posterior mean becomes more dominated by a single matrix satisfying the constraints in equation B.17 the higher c. No outcomes have as yet been observed for action 6, but at higher levels of control, its outcome is still inferred with high certainty due to the constraint that all actions lead to a different outcome. The four panels on the right show the effect of relaxing this assumption. **E** shows the data, which is the same as in A but rotated for clarity. **F** shows that for $c = 1/L$ the same predictive distribution is inferred. In comparison to C, **G** shows that a small value of c now really does lead to much uncertainty, as outcomes from different actions can no more be used to constrain each other "by elimination". **H** shows that for $c = 0.9$, low-entropy posteriors are only seen for those actions where outcomes have been observed. Note that for action 6, the posterior mean is flat.

prior

$$p(\mathbf{M}) = \frac{1}{L!}\left(\prod_j \delta(1-\sum_i M_{ij})\right)\left(\prod_i \delta(1-\sum_j M_{ij})\right)$$

to enforce the constraint that each row and column must contain one unit entry. Given a set of observations $\mathbf{N}$, this allows us to write the posterior distribution over $\mathbf{M}$ and the predictive distributions for action a as:

$$p(\mathbf{M}|\mathbf{N},c) = \frac{p(\mathbf{N}|c,\mathbf{M})p(\mathbf{M})}{p(\mathbf{N}|c)} \tag{B.17}$$

$$p(n_{D+1}=j|\mathbf{N},c,a) \propto \sum_{\mathbf{M}} c^{(\mathbf{m}^a)^\mathsf{T}\mathbf{d}^j}\bar{c}^{(1-\mathbf{m}^a)^\mathsf{T}\mathbf{d}^j}p(\mathbf{M}|\mathbf{N},c) \tag{B.18}$$

where $d_i^j = \delta_{ij}$. Figure B.2A shows the posterior mean $\mathbb{E}[\mathbf{M}|\mathbf{N},c]$ for three different values of c. As the c is shared between actions, and $\mathbf{M}$ assumes that all outcomes are achievable, this would mean that either, for $c \to 1$, all outcomes are achievable by precisely one action, or, for $c \to 1/L$, no outcome is controllably achievable. Figure B.2A-D illustrates the effects of such a constraint.

To relax this assumption, we allow the *number* $|M|$ of actions with controllably attainable outcomes to vary, i.e. each row and column of the matrix M can have either one unity entry, or none, as illustrated by $\mathbf{M}_2$ in equation B.16. Analogous to the previous section, we write a prior $p(|M|)$ over the set size $|M| = \sum_{ij} M_{ij} \leq L$ of controllably achievable outcomes and then integrate over it, leading to a prior over matrices

$$
\begin{aligned}
p(\mathbf{M}) &= \sum_{|M|=1}^{L} p(|M|) \left[\binom{L}{|M|} \frac{L!}{(L-|M|)!} \right]^{-1} B(\mathbf{M})\, \delta\left(\sum_{ij} M_{ij} - |M| \right) \qquad \text{(B.19)} \\
B(\mathbf{M}) &= \left(\prod_j \left[\delta\left(1 - \sum_i M_{ij} \right) + \delta\left(\sum_i M_{ij} \right) \right] \right) \times \\
&\quad \left(\prod_i \left[\delta\left(1 - \sum_j M_{ij} \right) + \delta\left(\sum_j M_{ij} \right) \right] \right)
\end{aligned}
$$

where $B(\mathbf{M})$ ensures that there is at most one unity entry in each row and column of the matrix M. In equation B.19 we let all matrices with the same number of entries have equal prior probability. For a matrix of size $L \times L$ with $|M| = k$, there are $\binom{L}{k}$ ways of choosing the columns, and $L!/(L-k)!$ was of filling the columns, as we care about the order.

In order to do prediction, we need to find the posterior distribution on the number of controllably achievable outcomes $|M|$, given the data, which is given by:

$$
p(|M| = k|\mathbf{n}, c) = \frac{\sum_{\mathbf{M}:|M|=k} p(\mathbf{N}|\mathbf{M}, c) p(\mathbf{M}|k)}{\sum_{|M|} p(|M|) \sum_{\mathbf{M}:|M|=k} p(\mathbf{N}|\mathbf{M}, c) p(\mathbf{M}|k)} \qquad \text{(B.20)}
$$

Thus if the prior $p(|M|) = \delta(|M| - L)$, we return to the previous setting where all outcomes have to be achievable if c is large enough. For priors that have mass on smaller $|M|$, not all outcomes have a dedicated action.

INFERENCE

For generalisation, it will necessary to infer the ML or MAP setting of c and $|M|$, given some data. This is again straightforward doing EM. For c we have:

$$
\begin{aligned}
\hat{c} &= \arg\max_c \log p(\mathbf{N}|c) \\
&= \arg\max_c \log \sum_{\mathbf{M}} p(\mathbf{N}|\mathbf{M}, c) p(M) \\
\Rightarrow \text{M step:} \quad 0 &= \left\langle \frac{\partial}{\partial c} \log p(\mathbf{M}, \mathbf{N}|c) \right\rangle_q = \left\langle \frac{\partial}{\partial c} \log p(\mathbf{N}|\mathbf{M}, c) \right\rangle_q \\
\Rightarrow \text{E step:} \quad q &\leftarrow p(\mathbf{M}|\mathbf{N}, c) \qquad \text{(B.21)}
\end{aligned}
$$

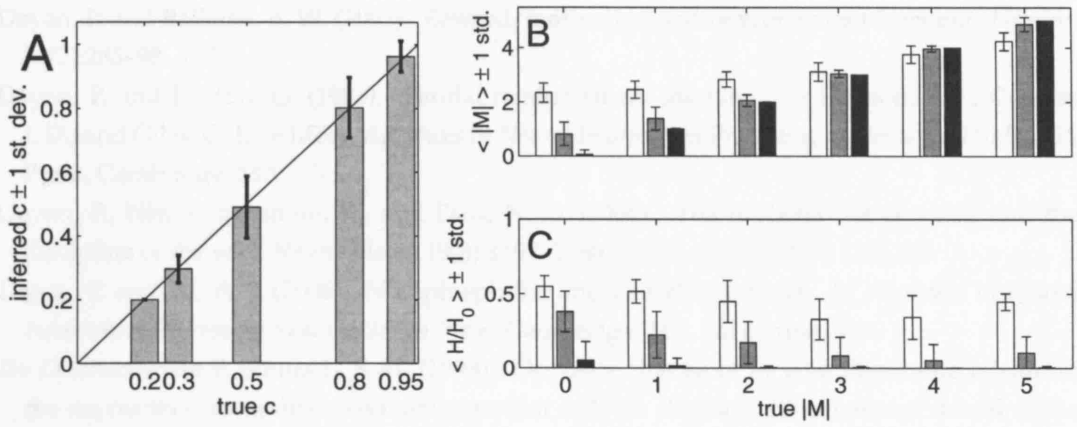


FIGURE B.3: **A**: Inference of c from outcome data by averaging over $p(\mathbf{M})$ as defined in equation B.19 using EM. For each inference, a total of 20 observations were obtained from a randomly chosen matrix $\mathbf{M}$ with the true underlying c, i.e. approx. 4 observations on each of $L = 5$ actions. **B**: Inferring $|M|$ from outcome data by equation B.23. White bars are for a total of 20 observations, grey bars for 50 and black bars for 100 observations. The black bars are very near the true values. $L = 5$ and $c = 0.9$. For small numbers of observations, the number of controllably achievable outcomes $|M|$ is overestimated, but with little confidence. **C**: Ratio of entropy of $p(k|\mathbf{N}, c)$ and a flat distribution with entropy $\mathcal{H} = -\log(1/N) \approx 1.6$

For small L, all the averages can be done explicitly. We evaluate the $p(\mathbf{N}|\mathbf{M}, c)$ for each of the $\mathbf{M}$ that have nonzero probability under our prior $p(\mathbf{M})$, and write

$$\begin{aligned} \left\langle \frac{\partial}{\partial c} \log p(\mathbf{N}|\mathbf{M}, c) \right\rangle_q &= \sum_{\mathbf{M}} \frac{\partial}{\partial c} \log p(\mathbf{N}|\mathbf{M}, c) w(\mathbf{M}) \\ w(\mathbf{M}) &= p(\mathbf{N}|\mathbf{M}, c) p(\mathbf{M}) \left(\sum_{\mathbf{M}} p(\mathbf{N}|\mathbf{M}, c) p(\mathbf{M}) \right)^{-1} \end{aligned} \tag{B.22}$$

Figure B.3A displays the characteristics of inference of c from data $\mathbf{N}$. Inference is very accurate. We will also look at the characteristics of generalisation based on $|M|$ and would thus like to infer it. We write for $|M| = k$:

$$p(k|N, c) \propto \sum_{\mathbf{M}:|\mathbf{M}|=k} p(\mathbf{N}|\mathbf{M}, c) p(\mathbf{M}|k) p(k) \tag{B.23}$$

which we maximise in the same way as we maximised the likelihood of c above. Figure B.3B and C show the performance of this inference. At small observation numbers, there is naturally little evidence for the low-control settings, and $|M|$ is overestimated.

Dayan, P. and Balleine, B. W. (2002). Reward, motivation and reinforcement learning. *Neuron*, 36(2):285–98. 127

Dayan, P. and Hinton, G. (1993). Feudal reinforcement learning. In Hanson, S. J., Cowan, J. D., and Giles, C. L., editors, *Advances in Neural Information Processing Systems (NIPS) 5*. MIT Press, Cambridge, MA. 83

Dayan, P., Niv, Y., Seymour, B., and Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw*, 19(8):1153–1160. 17, 18, 20, 108, 110, 120, 128

Dayan, P. and Yu, A. J. (2006). Norepinephrine and neural interrupts. In *Advances in Neural Information Processing Systems 18, In Press*, Cambridge, MA. MIT Press. 104

De Deurwaerdere P, Stinus L, S. U. (1998). Opposite change of in vivo dopamine release in the rat nucleus accumbens and striatum that follows electrical stimulationof dorsal raphe nucleus: role of 5-ht3 receptors. *J. Neurosci.*, 18(16):6528–38. 122

de Vasconcellos, A. P. S., Nieto, F. B., Fontella, F. U., da Rocha, E. R., and Dalmaz, C. (2006). The nociceptive response of stressed and lithium-treated rats is differently modulated by different flavors. *Physiol Behav*, 88(4-5):382–388. 54

Deakin, J. F. (2003). Depression and antisocial personality disorder: two contrasting disorders of 5ht function. *J Neural Transm Suppl*, 64:79–93. 33, 118, 121

Deakin, J. F. W. (1983). Roles of brain serotonergic neurons in escape, avoidance and other behaviors. *J Psychopharmacol*, 43:563–77. 110, 120

Deakin, J. F. W. and Graeff, F. G. (1991). 5-HT and mechanisms of defence. *Journal of Psychopharmacology*, 5:305–16. 19, 59, 82, 110, 120

Dearden, R., Friedman, N., and Andre, D. (1999). Model-based Bayesian exploration. In *Proceedings of the fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 150–9, Stockholm. 87, 104, 123, 138

Dearden, R., Friedman, N., and Russell, S. (1998). Bayesian Q-learning. In *Proceedings of the fifteenth National Conference on Artificial Intelligence*, pages 761–8. 64, 83, 87, 104, 123, 138

Deldin, J., Keller, J., Gergen, J. A., and Miller, G. A. (2001). Cognitive bias and emotion in neuropsychological models of depression. *Cogn Emotion*, 15(6):787–802. 29

Delgado, P. L. (2000). Depression: the case for a monoamine deficiency. *J Clin Psychiatry*, 61 Suppl 6:7–11. 33, 34

Delgado, P. L., Charney, D. S., Price, L. H., Aghajanian, G. K., Landis, H., and Heninger, G. R. (1990). Serotonin function and the mechanism of antidepressant action. reversal of antidepressant-induced remission by rapid depletion of plasma tryptophan. *Arch Gen Psychiatry*, 47(5):411–418. 33, 110, 121

Delgado, P. L., Price, L. H., Miller, H. L., Salomon, R. M., Aghajanian, G. K., Heninger, G. R., and Charney, D. S. (1994). Serotonin and the neurobiology of depression. effects of tryptophan depletion in drug-free depressed patients. *Arch Gen Psychiatry*, 51(11):865–874. 32, 33

DeMonbreun, B. G. and Craighead, W. E. (1977). Distortion of perception and recall of positive and neutral feedback in depression. *Cog Ther Res*, 1(4):311–29. 29

Depue, R. A. and Iacono, W. G. (1989). Neurobehavioral aspects of affective disorders. *Annu. Rev. Psychol.*, 40:457–92. 27

Depue, R. A. and Monroe, S. M. (1978). Learned helplessness in the perspective of the depressive disorders: conceptual and definitional issues. *J Abnorm Psychol*, 87(1):3–20. 62, 63

Dess, N. K., Linwick, D., Patterson, J., Overmier, J. B., and Levine, S. (1983). Immediate and

proactive effects of controllability and predictability on plasma cortisol responses to shocks in dogs. *Behav Neurosci*, 97(6):1005–1016. 84

Deutch, A. Y., Clark, W. A., and Roth, R. H. (1990). Prefrontal cortical dopamine depletion enhances the responsiveness of mesolimbic dopamine neurons to stress. *Brain Res.*, 521(1-2):311–5. 56

Di Chiara, G., Loddo, P., and Tanda, G. (1999). Reciprocal changes in prefrontal and limbic dopamine responsiveness to aversive and rewarding stimuli after chronic mild stress: implicationsfor the psychobiology of depression. *Biol Psychiatry*, 46(12):1624–33. 57, 83

Di Chiara, G. and Tanda, G. (1997). Blunting of reactivity of dopamine transmission to palatable food: a biochemical marker of anhedonia in the cms model? *Psychopharmacology (Berl)*, 134(4):351–3. 57

Dias, R., Robbins, T. W., and Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, 380:69–72. 43

Dickinson, A. and Balleine, B. (2002). The role of learning in the operation of motivational systems. In Gallistel, R., editor, *Stevens' handbook of experimental psychology*, volume 3, pages 497–534. Wiley, New York. 17, 61, 105, 110

Dickinson, A. and Dearing, M. F. (1979). Appetitive-aversive interactions and inhibitory processes. In Dickinson, A. and Boakes, R. A., editors, *Mechanisms of learning and motivation*, pages 203–231. Erlbaum, Hillsdale, NJ. 110

Dickinson, A., Shanks, D., and Evenden, J. (1984). Judgement of actoutcome contingency: The role of selective attribution. *Q J Exp Psych Human Exp Psych*, 36a:29–50. 38

Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *J Artif. Int. Res.*, 13:227–303. 66, 83

Doya, K. (2000). Metalearning, neuromodulation and emotion. In Hatano, G., Okada, N., and Ta, H., editors, *Affective minds*, pages 101–4. Elsevier Science, Amsterdam. 34, 122

Drevets, W. C., Price, J. L., Simpson, J. R., Todd, R. D., Reich, T., Vannier, M., and Raichle, M. E. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. *Nature*, 386:824–7. 17

Drugan, R. C., Ader, D. N., and Maier, S. F. (1985). Shock controllability and the nature of stress-induced analgesia. *Behav Neurosci*, 99(5):791–801. 63, 81

Drugan, R. C., Ryan, S. M., Minor, T. R., and Maier, S. F. (1984). Librium prevents the analgesia and shuttlebox escape deficit typically observed following inescapable shock. *Pharmacol. Biochem. Behav.*, 21(5):749–54. 55

Dulawa, S. C. and Hen, R. (2005). Recent advances in animal models of chronic antidepressant effects: the novelty-induced hypophagia test. *Neurosci Biobehav Rev*, 29(4-5):771–783. 52

Dworkin, R. H., Clark, W. C., and Lipsitz, J. D. (1995). Pain responsivity in major depression and bipolar disorder. *Psychiatry Res*, 56(2):173–181. 31

Ebstein, R. P., Benjamin, J., and Belmaker, R. H. (2000). Personality and polymorphisms of genes involved in aminergic neurotransmission. *Eur. J. Pharmacology*, 410(2-3):205–14. 107

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., Plomin, R., and Craig, I. W. (2004). Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Mol Psychiatry*, 9(10):908–915. 50

Elliott, R. (1998). The neuropsychological profile in unipolar depression. *Trends Cog. Sci.*, 2(11):447–54. 43

Elliott, R., Rubinsztein, J. S., Sahakian, B. J., and Dolan, R. J. (2002). The neural basis of mood-

congruent processing biases in depression. *Arch Gen Psychiatry*, 59(7):597–604. 29

Elliott, R., Sahakian, B. J., Herrod, J. J., Robbins, T. W., and Paykel, E. S. (1997). Abnormal response to negative feedback in unipolar depression: evidence for a diagnosis-specific impairment. *J. Neurol. Neurosurg. Psychiatry*, 63:74–82. 41

Elliott, R., Sahakian, B. J., McKay, A. P., Herrod, J. J., Robbins, T. W., and Paykel, E. S. (1996). Neuropsychological impairments in unipolar depression: the role of perceived failure on subsequent performance. *Psychol. Med.*, 26:975–89. 38, 41

Elliott, R., Sahakian, B. J., Michael, A., Paykel, E. S., and Dolan, R. J. (1998). Abnormal neural response to feedback on planning and guessing tasks in patients with unipolar depression. *Psychol Med*, 28(3):559–71. 18

Ellis, P. M. and Salmond, C. (1994). Is platelet imipramine binding reduced in depression? a meta-analysis. *Biol Psychiatry*, 36(5):292–299. 34

Engel, Y. (2005). *Algorithms and Representations for Reinforcement Learning*. PhD thesis, Hebrew University. 87

Erickson, K., Drevets, W. C., Clark, L., Cannon, D. M., Bain, E. E., Zarate, C. A., Charney, D. S., and Sahakian, B. J. (2005). Mood-congruent bias in affective go/no-go performance of unmedicated patients with major depressive disorder. *Am J Psychiatry*, 162(11):2171–2173. 29, 32

Esposito, E. (2006). Serotonin-dopamine interaction as a focus of novel antidepressant drugs. *Curr Drug Targets*, 7(2):177–185. 110

Estes, W. and Skinner, B. (1941). Some quantitative aspects of anxiety. *J. Exp. Psychol*, 29:390–400. 110

Everitt, B. J. and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci*, 8(11):1481–1489. 39

Eysenck, H. (1997). *Dimensions of Personality*. Transaction Publishers. 31, 50

Fallgatter, A. J., Herrmann, M. J., Roemmler, J., Ehlis, A. C., Wagener, A., Heidrich, A., Ortega, G., Zeng, Y., and Lesch, K. P. (2004). Allelic variation of serotonin transporter function modulates the brain electrical response for error processing. *Neuropsychopharm.*, 29(8):1506–11. 42

Fava, M. and Kendler, K. S. (2000). Major depressive disorder. *Neuron*, 28:335–41. 22, 24, 50

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., and Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry*, 26(1):57–63. 25

Fellows, L. K. and Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb. Cortex*, 15:58–63. 42

Fletcher, P. J. (1995). Effects of combined or separate 5,7-dihydroxytryptamine lesions of the dorsal and median raphe nuclei on respondingmaintained by a drl 20s schedule of food reinforcement. *Brain Res*, 675:45–54. 19, 59, 110

Fletcher, P. J. (1996). Injection of 5-ht into the nucleus accumbens reduces the effects of d-amphetamine on responding for conditioned reward. *Psychopharmacology (Berl.)*, 126(1):62–9. 18, 60, 82, 110

Fletcher, P. J., Azampanah, A., and Korth, K. M. (2002). Activation of 5-ht(1b) receptors in the nucleus accumbens reduces self-administration of amphetamine on a progressive ratio schedule. *Pharmacol Biochem Behav*, 71(4):717–21. 60

Fletcher, P. J. and Korth, K. M. (1999). Activation of 5-ht1b receptors in the nucleus accumbens reduces amphetamine-induced enhancement of responding for conditioned reward. *Psychopharmacology*, 142(2):165–74. 19, 82, 110, 122

Fletcher, P. J., Korth, K. M., and Chambers, J. W. (1999). Selective destruction of brain serotonin neurons by 5,7-dihydroxytryptamine increases responding for a conditioned reward. *Psychopharmacology (Berl).*, 147(3):291–9. 19, 60

Floresco, S. B., West, A. R., Ash, B., Moore, H., and Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci*, 6(9):968–973. 18, 122

Fontella, F. U., Nunes, M. L., Crema, L. M., Balk, R. S., Dalmaz, C., and Netto, C. A. (2004). Taste modulation of nociception differently affects chronically stressed rats. *Physiol Behav*, 80(4):557–561. 54, 108

Frank, M. J., Seeberger, L. C., and O'reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, 306(5703):1940–3. 18

Frazer, A. and Morilak, D. A. (2005). What should animal models of depression model? *Neurosci. Biobeh. Rev.*, 29:5150523. 28, 52, 62, 86

Friedhoff, A. J., Carr, K. D., Uysal, S., and Schweitzer, S. (1995). Repeated inescapable stress produces a neuroleptic-like effect on the conditioned avoidance response. *Neuropsychopharmacology*, 13(2):129–38. 54

Friedman, N. and Singer, Y. (1999). Efficient Bayesian Parameter Estimation in Large Discrete Domains. In Solla, S. A., Leen, T. K., and Müller, K.-R., editors, *Advances in Neural Information Processing Systems*, volume 11. MIT Press. 104, 138

Galina, Z.-H. and Amit, Z. (1986). Stress-induced analgesia: its effects on performance in learning paradigms. *Ann. NY Acad. Sci.*, 467:238–48. 81

Galloway, M. P., Suchowski, C. S., Keegan, M. J., and Hjorth, S. (1993). Local infusion of the selective 5ht-1b agonist cp-93,129 facilitates striatal dopamine release in vivo. *Synapse*, 15(1):90–2. 60, 122

Gambarana, C., Ghiglieri, O., and de Montis, M. G. (1995a). Desensitization of the d1 dopamine receptors in rats reproduces a model of escape deficit reverted by imipramine, fluoxetine and clomipramine. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, 19(5):741. 58

Gambarana, C., Ghiglieri, O., Tagliamonte, A., D'Alessandro, N., and de Montis, M. G. (1995b). Crucial role of D_1 dopamine receptors in mediating the antidepressant effect of imipramine. *Pharmacol Biochem Behav*, 50(2):147–51. 57

Gambarana, C., Masi, F., Leggio, B., Grappi, S., Nanni, G., Scheggi, S., De Montis, M. G., and Tagliamonte, A. (2003). Acquisition of a palatable-food-sustained appetitive behavior in satiated rats is dependent on the dopaminergic response to this food in limbic areas. *Neuroscience*, 121(1):179–87. 57

Gambarana, C., Masi, F., Tagliamonte, A., Scheggi, S., Ghiglieri, O., and De Montis, M. G. (1999). A chronic stress that impairs reactivity in rats also decreases dopaminergic transmission in the nucleus accumbens. *J. Neurochem.*, 72(5):2039–46. 55, 57, 105

Gambarana, C., Scheggi, S., Tagliamonte, A., Tolu, P., and Montis, M. G. D. (2001). Animal models for the study of antidepressant activity. *Brain Res Brain Res Protoc*, 7(1):11–20. 55

Ganesan, R. and Pearce, J. M. (1988). Effect of changing the unconditioned stimulus on appetitive blocking. *J Exp Psychol Anim Behav Process*, 14(3):280–291. 19

Gardner, J. and Oswald, A. (2001). Does money buy happiness? a longitudinal study using data on windfalls. Technical report, University of Warwick. http://www.nber.org/ confer/2001/midmf01/oswald.pdf. 105

Ghavamzadeh, M. and Engel, Y. (2007). Bayesian policy gradient algorithms. In *Proc. Advances in Neural Information Processing Systems (NIPS 19)*. 104

Ghiglieri, O., Gambarana, C., Scheggi, S., Tagliamonte, A., Willner, P., and De Montis, M. B. (1997). Palatable food induces an appetitive behaviour in satiated rats which can be inhibited by chronic stress. *Behav. Pharmacol.*, 8(6-7):619–28. 54, 55, 56, 57, 83, 108

Gibbon, J., Berryman, R., and Thompson, R. L. (1974). Contingency spaces and measures in classical and instrumental conditioning. *J Exp Anal Behav*, 21(3):585–605. 137

Gillespie, N. A., Whitfield, J. B., Williams, B., Heath, A. C., and Martin, N. G. (2005). The relationship between stressful life events, the serotonin transporter (5-httlpr) genotype and major depression. *Psychol Med*, 35(1):101–111. 50

Glass, D. C., Singer, J. E., Leonard, H. S., Krantz, D., Cohen, S., and Cummings, H. (1973). Perceived control of aversive stimulation and the reduction of stress responses. *J Pers*, 41(4):577–595. 106

Gloaguen, V., Cottraux, J., Cucherat, M., and ... (1998). A meta-analysis of the effects of cognitive therapy in depressed people. *J Affect Disord*, 49:59–72. 27

Gold, P. W., Drevets, W. C., and Charney, D. S. (2002). New insights into the role of cortisol and the glucocorticoid receptor in severe depression. *Biol Psychiatry*, 52(5):381–385. 30

Goldberg, T. E., Gold, J. M., Greenberg, R., Griffin, S., Schulz, S. C., Pickar, D., Kleinman, J. E., and Weinberger, D. R. (1993). Contrasts between patients with affective disorders and patients with schizophrenia on a neuropsychological test battery. *Am J Psychiatry*, 150(9):1355–1362. 42

Goldsmith, S. K., Pellmar, T. C., Kleinman, A. M., and Bunney, W. E., editors (2003). *Reducing suicide: a national imperative*. Institute of Medicine of the National Academies, The National Academies Press, Washington DC. 23

Golin, S., Terrell, F., and Johnson, B. (1977). Depression and the illusion of control. *J Abnorm Psychol*, 86(4):440–442. 37

Goodkin, F. (1976). Rats learn the relationship between responding and environmental events: An expansion of the learned helplessness hypothesis. *Learning and Motivation*, 7:382–393. 35, 53, 54, 84, 86, 105

Goodwin, F. K. and Jamison, K. R. (1990). *Manic-depressive Illness*. OUP. 44

Goodwin, G. M. (1997). Neuropsychological and neuroimaging evidence for the involvement of the frontal lobes in depression. *J Psychopharmacol*, 11(2):115–122. 38

Goto, Y. and Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci*, 8(6):805–812. 18, 122

Gottfried, J. A., O'Doherty, J., and Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, 301(5636):1104–1107. 17

Grabe, H. J., Lange, M., Wolff, B., Vlzke, H., Lucht, M., Freyberger, H. J., John, U., and Cascorbi, I. (2005). Mental and physical distress is modulated by a polymorphism in the 5-ht transporter gene interacting with social stressors and chronic disease burden. *Mol Psychiatry*, 10(2):220–224. 50

Graeff, F. G. (2002). On serotonin and experimental anxiety. *Psychopharmacology (Berl)*, 163(3-

4):467–476. 19, 59, 60, 82, 110

Graeff, F. G., Guimaraes, F. S., De Andrade, T. G. C. S., and Deakin, J. F. W. (1998). Role of 5HT in stress, anxiety and depression. *Pharm. Biochem. Behav.*, 54(1):129–41. 59, 82, 110, 127

Grahn, R. E., Maswood, S., McQueen, M. B., Watkins, L. R., and Maier, S. F. (1999a). Opioid-dependent effects of inescapable shock on escape behavior and conditioned fear responding are mediated by the dorsal raphe nucleus. *Behav Brain Res*, 99(2):153–167. 58

Grahn, R. E., Will, M. J., Hammack, S. E., Maswood, S., McQueen, M. B., Watkins, L. R., and Maier, S. F. (1999b). Activation of serotonin-immunoreactive cells in the dorsal raphe nucleus in rats exposed to an uncontrollable stressor. *Brain Res*, 826(1):35–43. 58

Grau, J. W., Hyson, R. L., Maier, S. F., Madden, J., and Barchas, J. D. (1981). Long-term stress-induced analgesia and activation of the opiate system. *Science*, 213:1409–11. 54, 81

Gray, J. A. (1991). *The psychology of fear and stress*, volume 5 of *Problems in the behavioural sciences*. Cambridge University Press, Cambridge, UK, 2 edition. 19, 32, 59, 82, 110, 120

Gray, J. A. and McNaughton, N. (2003). *The neuropsychology of anxiety*. OUP, 2nd edition. 82, 110, 111, 120

Greden, J. F., Genero, N., Price, H. L., Feinberg, M., and Levine, S. (1986). Facial electromyography in depression. subgroup differences. *Arch Gen Psychiatry*, 43(3):269–274. 29

Grippo, A. J., Beltz, T. G., and Johnson, A. K. (2003). Behavioral and cardiovascular changes in the chronic mild stress model of depression. *Physiol Behav*, 78(4-5):703–710. 55

Grippo, A. J., Sullivan, N. R., Damjanoska, K. J., Crane, J. W., Carrasco, G. A., Shi, J., Chen, Z., Garcia, F., Muma, N. A., and de Kar, L. D. V. (2005). Chronic mild stress induces behavioral and physiological changes, and may alter serotonin 1a receptor function, in male and cycling female rats. *Psychopharmacology (Berl)*, 179(4):769–780. 59

Haffel, G. J., Abramson, L. Y., Voelz, Z. R., Metalsky, G. I., Halberstadt, L., Dykman, B. M., Donovan, P., Hogan, M. E., Hankin, B. L., and Alloy, L. B. (2005). Negative cognitive styles, dysfunctional attitudes, and the remitted depression paradigm: a search for the elusive cognitive vulnerability to depression factor among remitted depressives. *Emotion*, 5(3):343–348. 51

Hale, A. S. (2005). The treatment of depression. In Griez, E. J. L., Faravelli, C., Nutt, D. J., and Zohar, J., editors, *Mood disorders: Clinical management and research issues*, chapter 10. John Wiley & Sons. 27

Hall, K. R. and Stride, E. (1954). The varying response to pain in psychiatric disorders: a study in abnormal psychology. *Br J Med Psychol*, 27(1-2):48–60. 31

Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry*, 23:56–62. 25

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci*, 26(32):8360–8367. 43

Harding, E. J., Paul, E. S., and Mendl, M. (2004). Animal behaviour: cognitive bias and affective state. *Nature*, 427:312. 56, 108

Hariri, A. R. and Holmes, A. (2006). Genetics of emotional regulation: the role of the serotonin transporter in neural function. *Trends Cog. Sci.*, 10(4):182–91. 33, 59, 121

Harrow, M., Yonan, C. A., Sands, J. R., and Marengo, J. (1994). Depression in schizophrenia: are neuroleptics, akinesia, or anhedonia involved? *Schizophr Bull*, 20(2):327–338. 47

Hasher, L. and Zacks, R. T. (1979). Automatic and effortful processes in memory. *J. Exp. Psychol.*

Gen., 108:356–88. 43

Hasler, G., Drevets, W. C., Manji, H. K., and Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology*, 29(10):1765–1781. 24, 26, 28

Hebb, D. O. (1955). Drives and the c.n.s. (conceptual nervous system). *Psychol Rev*, 62(4):243–254. 14

Heim, C., Newport, D. J., Heit, S., Graham, Y. P., Wilcox, M., Bonsall, R., Miller, A. H., and Nemeroff, C. B. (2000). Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *JAMA*, 284(5):592–7. 50

Heinz, A. (1999). Anhedonie – nosologieübergreifendes Korrelat einer Dysfunktion des dopaminergen Verstärkungssystem? *Nervenarzt*, 70(5):391–398. 27, 123

Henriques, J. B. and Davidson, R. J. (2000). Decreased responsiveness to reward in depression. *Cognition and Emotion*, 14(5):711–24. 40, 43

Henriques, J. B., Glowacki, J. M., and Davidson, R. J. (1994). Reward fails to alter response bias in depression. *J Abnorm Psychol*, 103(3):460–6. 17, 40, 43

Hershberger, W. A. (1986). An approach through the looking-glass. *Anim. Learn. Behav.*, 14:443–51. 18

Hettema, J. M., Kuhn, J. W., Prescott, C. A., and Kendler, K. S. (2006a). The impact of generalized anxiety disorder and stressful life events on risk for major depressive episodes. *Psychol Med*, 36(6):789–795. 24

Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A., and Kendler, K. S. (2006b). A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry*, 163(5):857–864. 107, 122

Hickie, I. (1996). Validity of the CORE: II: neuropsychological tests. In Parker, G. and Hadzi-Pavlovic, D., editors, *Melancholia: a disorder of movement and mood*, chapter 9, pages 149–59. Cambridge University Press. 44

Hickie, I., Parsonage, B., and Parker, G. (1990). Prediction of response to electroconvulsive therapy. preliminary validation of a sign-based typology of depression. *Br. J. Psychiatry*, 157:65–71. 44

Higgins, G. A. and Fletcher, P. J. (2003). Serotonin and drug reward: focus on 5-ht2c receptors. *Eur J Pharmacol*, 480(1-3):151–162. 19, 60

Hiroto, D. and Seligman, M. (1975). Generality of learned helplessness in man. *Journal of Personality and Social Psychology*, 31(2):311–327. 106

Holland, P. C. (1979). Differential effects of omission contingencies on various components of Pavlovian appetitive conditioned responding in rats. *J. Exp. Psych. Animal Beh. Proc.*, 5(2):178–93. 18

Hollerman, J. R., Tremblay, L., and Schultz, W. (2000). Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. *Prog Brain Res*, 126:193–215. 17

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96(4):651–6. 56, 104

Huesmann, L. R. (1978). Cognitive processes and models of depression. *J Abnorm Psychol*, 87(1):194–198. 36

Hughes, J. R., Pleasants, C. N., and Pickens, R. W. (1985). Measurement of reinforcement in depression: a pilot study. *J Behav Ther Exp Psychiatry*, 16(3):231–6. 45, 46

Hull, C. (1943). *Principles of behavior*. Appleton-Century-Crofts New York. 28

Huys, Q. J. M. (2006). Smooth priors for point process observations. Minor project proposal. Gatsby Computational Neuroscience Unit. UCL. 11

Huys, Q. J. M., Ahrens, M. B., and Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *J Neurophysiol*, 96(2):872–890. 11

Huys, Q. J. M. and Paninski, L. (2006). Model-based optimal interpolation and filtering for noisy, intermittent biophysical recordings. *Fifteenth Annual Computational Neuroscience Meeting*. 11

Huys, Q. J. M., Zemel, R., Natarajan, R., and Dayan, P. (2007). Fast population coding. *Neural Computation.*, 19(2):460–97. 11

Imperato, A., Cabib, S., and Puglisi-Allegra, S. (1993). Repeated stressful experiences differently affect the time-dependent responses of the mesolimbic dopamine system to the stressor. *Brain Res.*, 601(1-2):333–6. 57

Imperato, A., Puglisi-Allegra, S., Casolini, P., and Angelucci, L. (1991). Changes in brain dopamine and acetylcholine release during and following stress are independent of the pituitary-adrenocortical axis. *Brain Res*, 538:111–7. 56, 57

Ivanov-Smolensky, A. G. (1925). Ueber die bedingten Reflexe in der depressiven Phase des manisch depressiven Irreseins. *Monatsschrift für Psychiatrie und Neurologie*, 58:376. 39

Iversen, L. (2005). The monoamine hypothesis of depression. In Licinio, J. and Wong, M.-L., editors, *Biology of depression*, volume 1, pages 71–86. Wiley, Weinheim, Germany. 33, 34

Jackson, R. L., Alexander, J. H., and Maier, S. F. (1980). Learned helplessness, inactivity and associative deficits: Effects of inescapable shock on response choice escape learning. *J. Exp. Psychol. Animal Behav. Proc.*, 6:1–20. 54

Jackson, R. L., Maier, S. F., and Coon, D. J. (1979). Long-term analgesic effects of inescapable shock and learned helplessness. *Science*, 206(4414):91–4. 54, 63, 81

Jackson, R. L., Maier, S. F., and Rapaport, P. M. (1978). Exposure to inescapable shock produces both activity and associative deficits in the rat. *Learn. Motiv.*, 9:69–98. 7, 53, 63, 64, 65, 67, 71, 73, 74, 99, 108

Jacobs, B. L. and Fornal, C. A. (1997). Serotonin and motor activity. *Curr. Op. Neurobiol.*, 7:820–5. 19

Jamison, K. R. (1997). *An Unquiet Mind*. Vintage. 14

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK. 103

Job, R. F. S. (2002). The effects of uncontrollable, unpredictable aversive and appetitive events: similar effects warrant similar, but not identical, explanations? *Integr Physiol Behav Sci*, 37(1):59–81. 54, 105, 106

Judd, L. L., Akiskal, H. S., Maser, J. D., Zeller, P. J., Endicott, J., Coryell, W., Paulus, M. P., Kunovac, J. L., Leon, A. C., Mueller, T. I., Rice, J. A., and Keller, M. B. (1998). Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. *J Affect Disord*, 50(2-3):97–108. 24

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292. 82

Kapur, S. and Mann, J. J. (1992). Role of the dopaminergic system in depression. *Biol. Psychiatry*, 32(1):1–17. 18, 27, 47

Kapur, S. and Remington, G. (1996). Serotonin-dopamine interaction and its relevance to

schizophrenia. *Am J Psychiatry*, 153(4):466–76. 60, 82, 110, 122

Kapur, S. and Remington, G. (2001). Atypical antipsychotics: new directions and new challenges in the treatment of schizophrenia. *Annu Rev Med*, 52:503–517. 60

Kapur, S. and Seeman, P. (2001). Does fast dissociation from the dopamine d(2) receptor explain the action of atypical antipsychotics?: A new hypothesis. *Am J Psychiatry*, 158(3):360–369. 60

Kasper, S., Boer, J., and Sitsen, J. A., editors (2003). *Handbook of Depression and Anxiety*. Marcel Dekker, New York, Basel, second edition. 49, 50

Katona, C., Peveler, R., Dowrick, C., Wessely, S., Feinmann, C., Gask, L., Lloyd, H., Williams, A. C. C., and Wager, E. (2005). Pain symptoms in depression: definition and clinical significance. *Clin. Med.*, 5(4):3990–5. 31

Kaufman, J. and Charney, D. (2000). Comorbidity of mood and anxiety disorders. *Depression Anxiety*, 12(S1):69–76. 23, 24, 25, 107, 122

Kaufman, J., Yang, B.-Z., Douglas-Palumberi, H., Houshyar, S., Lipschitz, D., Krystal, J. H., and Gelernter, J. (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proc Natl Acad Sci U S A*, 101(49):17316–17321. 50

Kearns, M. and Singh, S. (1998). Near-optimal reinforcement learning in polynomial time. In *Proc.15th Int. Conf. Machine Learning*. Morgan Kaufmann. 104

Keller, M. B., Klerman, G. L., Lavori, P. W., Coryell, W., Endicott, J., and Taylor, J. (1984). Long-term outcome of episodes of major depression: Clinical and public health significance. *J Am Med Ass*, 252:788–92. 25

Kelly, D. D., editor (1986). *Stress-induced analgesia*. N Y Acad. Med. Sci. 54, 63

Kendler, K. S., Hettema, J. M., Butera, F., Gardner, C. O., and Prescott, C. A. (2003a). Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. *Arch Gen Psychiatry*, 60(8):789–796. 107

Kendler, K. S., Karkowski, L. M., and Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry*, 156:837–41. 50, 107

Kendler, K. S., Kessler, R. C., Walters, E. E., MacLean, C., Neale, M. C., Heath, A. C., and Eaves, L. J. (1995). Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am J Psychiatry*, 152(6):833–842. 50

Kendler, K. S., Kuhn, J. W., Vittum, J., Prescott, C. A., and Riley, B. (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Arch Gen Psychiatry*, 62(5):529–535. 50

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., and Eaves, L. J. (1992). Major depression and generalized anxiety disorder. same genes, (partly) different environments? *Arch Gen Psychiatry*, 49(9):716–722. 23

Kendler, K. S., Prescott, C. A., Myers, J., and Neale, M. C. (2003b). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry*, 60(9):929–937. 25, 26, 107

Kendler, K. S., Thornton, L. M., and Gardner, C. O. (2000). Stressful life events and previous episodes in the etiology of major depression in women: An evaluation of the "kindling" hypothesis. *Am. J. Psychiatry*, 157:1243–51. 50, 107

Kessler, R. C. (1997). The effects of stressful life events on depression. *Annu Rev Psychol*, 48:191–214. 23, 49, 107

Killcross, S. and Coutureau, E. (2003). Coordination of actions and habits in the medial pre-

frontal cortex of rats. *Cereb Cortex*, 13(4):400–408. 17, 54, 127

Klaassen, T., Klumperbeek, J., Deutz, N. E., van Praag, H. M., and Griez, E. (1998). Effects of tryptophan depletion on anxiety and on panic provoked by carbon dioxide challenge. *Psychiatry Res*, 77(3):167–174. 35, 121

Klaassen, T., Riedel, W. J., Deutz, N. E., and Van Praag, H. M. (2002). Mood congruent memory bias induced by tryptophan depletion. *Psychol. Med.*, 32(1):167–72. 34, 116, 121

Kleinman, A. (2004). Culture and depression. *N Engl J Med*, 351(10):951–953. 24, 25

Knuth, D. and Moore, R. (1975). An Analysis of Alpha-Beta Pruning. *Artificial Intelligence*, 6(4):293–326. 110, 120

Koolhaas, J. M., Korte, S. M., Boer, S. F. D., Vegt, B. J. V. D., Reenen, C. G. V., Hopster, H., Jong, I. C. D., Ruis, M. A. W., and Blokhuis, H. J. (1999). Coping styles in animals: current status in behavior and stress-physiology. *Neurosci. Biobehav. Rev.*, 23(7):925–35. 30

Korf, J. and van Praag, H. M. (1971a). Endogenous depressions with and without disturbances in the 5-hydroxytryptamine metabolism: A biochemical classification? *Psychopharmacologia*, 19(2):148–152. 33

Korf, J. and van Praag, H. M. (1971b). Retarded depression and the dopamine metabolism. *Psychopharmacologia*, 19(2):199–203. 46

Korte, S. M., Koolhaas, J. M., Wingfield, J. C., and McEwen, B. S. (2005). The darwinian concept of stress: benefits of allostasis and costs of allostatic load and the trade-offs in health and disease. *Neurosci Biobehav Rev*, 29(1):3–38. 30

Kutchins, H. and Kirk, S. A. (1997). *Making us crazy: DSM – the psychiatric bible and the creation of mental disorders*. Free Press, New York. 25

Laaris, N., Poul, E. L., Laporte, A. M., Hamon, M., and Lanfumey, L. (1999). Differential effects of stress on presynaptic and postsynaptic 5-hydroxytryptamine-1a receptors in the rat brain: an in vitroelectrophysiological study. *Neuroscience*, 91(3):947–58. 59

Lanfumey, L., Pardon, M. C., Laaris, N., Joubert, C., Hanoun, N., Hamon, M., and Cohen-Salmon, C. (1999). 5-HT_{1A} autoreceptor desensitization by chronic ultramild stress in mice. *Neuroreport*, 10(16):3369–74. 59

Lapin, I. P. and Oxenkrug, G. F. (1969). Intensification of the central serotoninergic processes as a possible determinant of the thymoleptic effect. *Lancet*, 1(7586):132–136. 27, 32, 58

Lautenbacher, S., Roscher, S., Strian, D., Fassbender, K., Krumrey, K., , and Krieg, J.-C. (1994). Pain perception in depression: Relationships to symptomatology and naloxone-sensitive mechanisms. *Psychosom. Med.*, 56:345–52. 31

Lautenbacher, S., Spernal, J., Schreiber, W., and Krieg, J. C. (1999). Relationship between clinical pain complaints and pain sensitivity in patients with depression and panic disorder. *Psychosom Med*, 61(6):822–827. 31

Layne, C. (1980). Motivational deficit in depression: people's expectations x outcomes' impacts. *J Clin Psychol*, 36(3):647–652. 17, 28, 43, 46

Lechin, F., Van Der Dijs, B., and Benaim, M. (1996). Stress versus depression. *Prog. Neuro-Psychopharm. Biol. Psych.*, 20(6):899–950. 30

Lee, C. and Rodgers, R. J. (1990). Antinociceptive effects of elevated plus-maze exposure: influence of opiate receptor manipulations. *Psychopharmacology*, 102(4):507–13. 54, 56

Lee, R. K. and Maier, S. F. (1988). Inescapable shock and attention to internal versus external cues in a water discrimination escape task. *J. Exp. Psychol. Anim. Behav. Process.*, 14(3):302–10.

53

LeMarquand, D. G., Benkelfat, C., Pihl, R. O., Palmour, R. M., and Young, S. N. (1999). Behavioral disinhibition induced by tryptophan depletion in nonalcoholic young men with multigenerational family histories of paternal alcoholism. *Am J Psychiatry*, 156(11):1771–1779. 121

Lemelin, S. and Baruch, P. (1998). Clinical psychomotor retardation and attention in depression. *J. Psychiatr. Res.*, 32(2):81–8. 44, 47, 48

Lengyel, M. and Dayan, P. (2007). Hippocampal contributions to control: a normative perspective. In *Comp. Sys. Neurosci.* 17, 108

Leonardo, E. D. and Hen, R. (2006). Genetics of affective and anxiety disorders. *Annu Rev Psychol*, 57:117–137. 50

Lesch, K.-P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., andClemens R Müller, J. B., Hamer, D. H., and Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, 274(5292):1527–31. 34, 59, 121

Lewinsohn, P., Youngren, M., and Grosscup, S. (1979). Reinforcement and depression. In Depue, R. A., editor, *The psychobiology of depressive disorders: Implications for the effects of stress*, pages 291–316. Academic Press, New York. 14, 17, 28, 30, 31, 105

Lewinsohn, P. M., Allen, N. B., Seeley, J. R., and Gotlib, I. H. (1999). First onset versus recurrence of depression: differential processes of psychosocial risk. *J Abnorm Psychol*, 108(3):483–489. 24

Licinio, J. and Wong, M.-L., editors (2004). *The biology of depression*. Wiley. 14

Lira, A., Zhou, M., Castanon, N., Ansorge, M. S., Gordon, J. A., Francis, J. H., Bradley-Moore, M., Lira, J., Underwood, M. D., Arango, V., Kung, H. F., Hofer, M. A., Hen, R., and Gingrich, J. A. (2003). Altered depression-related behaviors and functional changes in the dorsal raphe nucleus of serotonin transporter-deficient mice. *Biol Psychiatry*, 54(10):960–971. 59

Little, K. Y. (1988). Amphetamine, but not methylphenidate, predicts antidepressant efficacy. *J Clin Psychopharmacol*, 8(3):177–183. 47

Lotrich, F. E., Pollock, B. G., and Ferrell, R. E. (2001). Polymorphism of the serotonin transporter: implications for the use of selective serotonin reuptake inhibitors. *Am J Pharmacogenomics*, 1(3):153–164. 32

Lyon, H. M., Startup, M., and Bentall, R. P. (1999). Social cognition and the manic defense: attributions, selective attention, and self-schema in bipolar affective disorder. *J Abnorm Psychol*, 108(2):273–282. 37, 51

Lysle, D. T. and Fowler, H. (1988). Changes in pain reactivity induced by unconditioned and conditioned excitatory and inhibitory stimuli. *J. Exp. Psychol. Anim. Behav. Process.*, 14(4):376–89. 54, 84

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge, UK. 87

Mahadevan, S. (1996). Average reward reinforcement learning: foundations, algorithms and results. *Machine Learning*, 22:159–95. 18, 67, 135

Maia, T. V. and McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: what participants really know in the iowa gambling task. *Proc Natl Acad Sci U S A*, 101(45):16075–16080. 42

Maier, S. and Seligman, M. (1976). Learned Helplessness: Theory and Evidence. *Journal of*

Experimental Psychology: General, 105(1):3–46. 17, 35, 53, 62, 63, 66, 86, 87, 89, 106, 107, 137

Maier, S. F. (1989). Determinants of the nature of environmentally induced hypoalgesia. *Behav. Neurosci.*, 103(1):131–43. 54, 63, 64, 65, 81

Maier, S. F. (2001). Exposure to the stressor environment prevents the temporal dissipation of behavioral depression/learned helplessness. *Biol Psychiatry*, 49(9):763–773. 55, 56

Maier, S. F., Amat, J., Baratta, M. V., Paul, E., and Watkins, L. R. (2006). Behavioral control, the medial prefrontal cortex, and resilience. *Dialogues Clin Neurosci*, 8(4):397–406. 17, 51

Maier, S. F., Busch, C. R., Maswood, S., Grahn, R. E., and Watkins, L. R. (1995a). The dorsal raphe nucleus is a site of action mediating the behavioral effects of the benzodiazepine receptor inverse agonist dmcm. *Behav. Neurosci.*, 109(4):759–66. 58

Maier, S. F., Drugan, R. C., and Grau, J. W. (1982). Controllability, coping behavior and stress-induced analgesia in the rat. *Pain*, 12:47–56. 63

Maier, S. F., Grahn, R. E., Kalman, Sutton, Wiertelak, and Watkins, L. R. (1993). Role of amygdala and dorsal raphe nucleus in mediating the behavioural consequences of inescapable shock. *Behav. Neurosci*, 107:377–88. 55, 58, 82

Maier, S. F., Grahn, R. E., and Watkins, L. R. (1995b). 8-OH-DPAT microinjected in the region of the dorsal raphe nuleus blocks and reverses the enhancement of fear conditioning and interference with escape produced by exposure to inescapable shock. *Behav. Neurosci.*, 109(3):404–12. 58, 82

Maier, S. F. and Watkins, L. R. (1998). Stressor controllability, anxiety, and serotonin. *Cog. Ther. Res.*, 22(6):595–613. 54, 58, 59

Maier, S. F. and Watkins, L. R. (2005). Stressor controllability and learned helplessness: the roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. *Neurosci. Biobehav. Rev.*, 29(4-5):829–41. 19, 52, 53, 54, 55, 58, 59, 62, 63, 64, 66, 77, 82, 86, 99, 108, 110

Malison, R. T., Price, L. H., Berman, R., van Dyck, C. H., Pelton, G. H., Carpenter, L., Sanacora, G., Owens, M. J., Nemeroff, C. B., Rajeevan, N., Baldwin, R. M., Seibyl, J. P., Innis, R. B., and Charney, D. S. (1998). Reduced brain serotonin transporter availability in major depression as measured by [123i]-2 beta-carbomethoxy-3 beta-(4-iodophenyl)tropane and single photon emission computed tomography. *Biol Psychiatry*, 44(11):1090–1098. 34

Mangiavacchi, S., Masi, F., Scheggi, S., Leggio, B., De Montis, M. G., and Gambarana, C. (2001). Long-term behavioral and neurochemical effects of chronic stress exposure in rats. *J. Neurochem.*, 79(6):2113–21. 54, 83

Manicavasagar, V., Silove, D., and Hadzi-Pavlovic, D. (1998). Subpopulations of early separation anxiety: relevance to risk of adult anxiety disorders. *J Affect Disord*, 48(2-3):181–190. 23

Mann, J. J. (1999). Role of the serotonergic system in the pathogenesis of major depression and suicidal behavior. *Neuropsychopharmacology*, 21(2 Suppl):99S–105S. 27, 32, 33

Mann, J. J., Huang, Y. Y., Underwood, M. D., Kassir, S. A., Oppenheim, S., Kelly, T. M., Dwork, A. J., and Arango, V. (2000). A serotonin transporter gene promoter polymorphism (5-httlpr) and prefrontal cortical binding in major depression and suicide. *Arch Gen Psychiatry*, 57(8):729–738. 34

Marr, D. (1982). *Vision*. Freeman, New York, NY, USA. 19

Martin, I. and Rees, L. (1966). Reaction times and somatic reactivity in depressed patients. *J Psychosom Res*, 9(4):375–382. 39

Masi, F., Scheggi, S., Mangiavacchi, S., Tolu, P., Tagliamonte, A., De Montis, M. G., and Gambarana, C. (2001). Dopamine output in the nucleus accumbens shell is related to the acquisition and the retention of a motivated appetitive behavior in rats. *Brain Res.*, 903(1-2):102–9. 57

Maudhuit, C., Prvot, E., Dangoumau, L., Martin, P., Hamon, M., and Adrien, J. (1997). Antidepressant treatment in helpless rats: effect on the electrophysiological activity of raphe dorsalis serotonergic neurons. *Psychopharmacology (Berl)*, 130(3):269–275. 32

Mayberg, H. S. (1997). Limbic-cortical dysregulation: a proposed model of depression. *J Neuropsych.*, 9(3):471–81. 17, 26

McAllister-Williams, R. H. and Tyrer, S. P. (2003). Antidepressants for the treatment of depression and anxiety disorders: same mechanism of action? In Kasper, S., den Boer, J. A., and Sitsen, J. M. A., editors, *Handbook of Depression and Anxiety*, chapter 18, pages 443–56. Marcel Dekker, second edition. 59

McClure, S. M., Berns, G. S., and Montague, P. R. (2003a). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2):339–46. 17

McClure, S. M., Daw, N. D., and Montague, P. R. (2003b). A computational substrate for incentive salience. *TINS*, 26:423–8. 17

McCutcheon, N. B., Rosellini, R. A., and Bandel, S. (1991). Controllability of stressors and rewarding brain stimulation: effect on the rate-intensity function. *Physiol Behav*, 50(1):161–166. 54, 83

McKinney, W. T. and Bunney, W. E. (1969). Animal model of depression. i. review of evidence: implications for research. *Arch Gen Psychiatry*, 21(2):240–248. 52

McNaughton, N. and Corr, P. J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neurosci Biobehav Rev*, 28(3):285–305. 110, 120, 121

McTavish, S. F. B., Mannie, Z. N., Harmer, C. J., and Cowen, P. J. (2005). Lack of effect of tyrosine depletion on mood in recovered depressed women. *Neuropsychopharmacology*, 30(4):786–91. 47

Mentis, M. J. and Delalot, D. (2005). Depression in Parkinson's disease. *Adv. Neurol.*, 42:26–41. 27, 47

Meyer, J. H., Krger, S., Wilson, A. A., Christensen, B. K., Goulding, V. S., Schaffer, A., Minifie, C., Houle, S., Hussey, D., and Kennedy, S. H. (2001). Lower dopamine transporter binding potential in striatum during depression. *Neuroreport*, 12(18):4121–4125. 46

Mikulincer, M. (1988). Reactance and helplessness following exposure to unsolvable problems: the effects of attributional style. *J Pers Soc Psychol*, 54(4):679–686. 106

Mikulincer, M. (1994). *Human learned helplessness: a coping perspective.* The Plenum Series in social/clinical psychology. Plenum Press, New York. 106

Millan, M. J. (2003). The neurobiology and control of anxious states. *Prog Neurobiol*, 70(2):83–244. 59

Millan, M. J. (2006). Multi-target strategies for the improved treatment of depressive states: Conceptual foundations and neuronal substrates, drug discovery and therapeutic application. *Pharmacol Ther*, 110(2):135–370. 32, 33, 47

Miller, H. E., Deakin, J. F., and Anderson, I. M. (2000). Effect of acute tryptophan depletion on co2-induced anxiety in patients with panic disorder and normal volunteers. *Br J Psychiatry*, 176:182–188. 35, 121

Miller, H. L., Delgado, P. L., Salomon, R. M., Berman, R., Krystal, J. H., Heninger, G. R., and Charney, D. S. (1996a). Clinical and biochemical effects of catecholamine depletion on antidepressant-induced remission of depression. *Arch Gen Psychiatry*, 53(2):117–128. 33

Miller, H. L., Delgado, P. L., Salomon, R. M., Heninger, G. R., and Charney, D. S. (1996b). Effects of alpha-methyl-para-tyrosine (ampt) in drug-free depressed patients. *Neuropsychopharmacology*, 14(3):151–157. 33

Miller, S. M. (1979). Controllability and human stress: method, evidence and theory. *Behav Res Ther*, 17(4):287–304. 63, 86, 106

Miller, W. R. (1975). Psychological deficit in depression. *Psychol Bull*, 82(2):238–260. 43, 47

Miller, W. R. and Seligman, M. E. (1975). Depression and learned helplessness in man. *J Abnorm Psychol*, 84(3):228–238. 35, 36, 86, 106

Miller, W. R., Seligman, M. E., and Kurlander, H. M. (1975). Learned helplessness, depression, and anxiety. *J Nerv Ment Dis*, 161(5):347–357. 40

Mineka, S. and Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Ann. Rev. Psychol.*, 36:495–529. 53, 84, 105, 106

Mineka, S., Watson, D., and Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annu Rev Psychol*, 49:377–412. 23, 24, 107

Minor, T. and LoLordo, V. (1984). Escape deficits following inescapable shock: The role of contextual odor. *Journal of Experimental Psychology: Animal Behavior Processes*, 10:168–181. 53

Minor, T. R., Jackson, R. L., and Maier, S. F. (1984). Effects of task irrelevant cues and reinforcement delay on choice escape learning following inescapable shock: The effects of a feedback stimulus. *J. Exp. Psychol. Animal. Beh. Proc.*, 10:168–81. 54

Mirenowicz, J. and Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, 379:449–51. 56

Mitra, R., Jadhav, S., McEwen, B. S., Vyas, A., and Chattarji, S. (2005). Stress duration modulates the spatiotemporal patterns of spine formation in the basolateral amygdala. *Proc Natl Acad Sci U S A*, 102(26):9371–9376. 108

Mobini, S., Chiang, T. J., Al-Ruwaitea, A. S., Ho, M. Y., Bradshaw, C. M., and Szabadi, E. (2000a). Effect of central 5-hydroxytryptamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology*, 149(3):313–8. 59, 122

Mobini, S., Chiang, T. J., Ho, M. Y., Bradshaw, C. M., and Szabadi, E. (2000b). Effects of central 5-hydroxytryptamine depletion on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology*, 152(4):390–7. 122

Mogil, J. S., Sternberg, W. F., Balian, H., Liebeskind, J. C., and Sadowski, B. (1996). Opioid and nonopioid swim stress-induced analgesia: A parametric analysis in mice. *Physiol. Behav.*, 59(1):123–32. 56

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J. Neurosci.*, 16(5):1936–47. 17, 82, 104, 122

Moreau, J.-L., Jenck, F., Martin, J. R., Mortas, P., and Haefely, W. E. (1992). Antidepressant treatment prevents chronic unpredictable mild stress-induced anhedonia as assessed by ventral tegmentum self-stimulation behavior in rats. *Eur. Neuropsychopharmacology*, 2(1):43–49. 55, 57, 83

Moreno, F. A., Gelenberg, A. J., Heninger, G. R., Potter, R. L., McKnight, K. M., Allen, J., Phillips,

A. P., and Delgado, P. L. (1999). Tryptophan depletion and depressive vulnerability. *Biol Psychiatry*, 46(4):498–505. 33, 110, 121

Msetfi, R. M., Murphy, R. A., Simpson, J., and Kornbrot, D. E. (2005). Depressive realism and outcome density bias in contingency judgments: the effect of the context and intertrial interval. *J. Exp. Psychol. Gen.*, 134(1):10–22. 38, 87

Mulrow, C. D., Williams, J. W., Chiqueete, E., and ... (2000). Efficacy of newer medications for treating depression in primary care people. *Am Med J*, 108:54–64. 27

Murphy, D. L., Li, Q., Engel, S., Wichems, C., Andrews, A., Lesch, K. P., and Uhl, G. (2001a). Genetic perspectives on the serotonin transporter. *Brain Res Bull*, 56(5):487–494. 50

Murphy, F. C., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., Paykel, E. S., and Sahakian, B. J. (2001b). Decision-making cognition in mania and depression. *Psychol. Med.*, 31:679–93. 38

Murphy, F. C., Sahakian, B. J., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., and Paykel, E. S. (1999). Emotional bias and inhibitory control processes in mania and depression. *Psych. Med.*, 29:1307–21. 29, 32, 43

Murphy, F. C., Smith, K. A., Cowen, P. J., Robbins, T. W., and Sahakian, B. J. (2002). The effects of tryptophan depletion on cognitive and affective processing in healthy volunteers. *Psychopharm.*, 163:42–53. 34, 121, 122

Murua, V. S., Gomez, R. A., Andrea, M. E., and Molina, V. A. (1991). Shuttle-box deficits induced by chronic variable stress: Reversal by imipramine administration. *Pharmacol. Biochem. Behav.*, 38(1):125–30. 55

Muscat, R., Papp, M., and Willner, P. (1992). Reversal of stress-induced anhedonia by the atypical antidepressants, fluoxetine and maprotiline. *Psychopharmacology (Berl)*, 109(4):433–438. 59

Muscat, R. and Willner, P. (1992). Suppression of sucrose drinking by chronic mild unpredictable stress: a methodological analysis. *Neurosci Biobehav Rev*, 16(4):507–517. 56

Must, A., Szabo, Z., Bodi, N., Szasz, A., Janka, Z., and Keri, S. (2006). Sensitivity to reward and punishment and the prefrontal cortex in major depression. *J. Affect. Disord.*, 90(2-3):209–15. 38, 42

Myin-Germeys, I., Peeters, F., Havermans, R., Nicolson, N. A., DeVries, M. W., Delespaul, P., and Os, J. V. (2003). Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr Scand*, 107(2):124–131. 29, 30, 107

Naber, D. (1988). Clinical relevance of endorphins in psychiatry. *Prog. Neuropsychopharmacol Biol. Psychiatry.*, 12(S1):119–135. 31

Naismith, S. L., Hickie, I. B., Ward, P. B., Scott, E., and Little, C. (2006). Impaired implicit sequence learning in depression: a probe for frontostriatal dysfunction? *Psychol. Med.*, 36(3):313–23. 47

Nanni, G., Scheggi, S., Leggio, B., Grappi, S., Masi, F., Rauggi, R., and De Montis, M. G. (2003). Acquisition of an appetitive behavior prevents development of stress-induced neurochemical modifications in rat nucleus accumbens. *J. Neurosci. Res.*, 73(4):573–80. 83

Naranjo, C. A., Tremblay, L. K., and Busto, U. E. (2001). The role of the brain reward system in depression. *Prog. Neuro-Psychopharmacol.*, 25:781–825. 27, 57

Natarajan, R., Huys, Q. J., Dayan, P., and Zemel, R. S. (2007). Online learning and inference in spiking populations. In preparation. 11

Nelson, A. and Killcross, S. (2006). Amphetamine exposure enhances habit formation. *J Neurosci*, 26(14):3805–3812. 17

Nelson, J. C. and Charney, D. S. (1981). The symptoms of major depressive illness. *Am J Psychiatry*, 138(1):1–13. 25, 44

Nelson, R. E. and Craighead, W. E. (1977). Selective recall of positive and negative feedback, self-control behaviors, and depression. *J Abnorm Psychol*, 86(4):379–388. 29

Nesse, R. M. (2000). Is depression and adaptation? *Arch. Gen. Psychiatry*, 57:14–20. 30, 103, 127

Neumeister, A., Konstantinidis, A., Stastny, J., Schwarz, M. J., Vitouch, O., Willeit, M., Praschak-Rieder, N., Zach, J., de Zwaanand B Bondy, M., Ackenheil, M., and Kasper, S. (2002). Association between serotonin transporter gene promoter polymorphism (5httlpr) and behavioral responses to tryptophan depletion in healthywomen with and without family history of depression. *Arch. Gen. Psychiatry*, 59(7):613–20. 33, 121

Ninan, P. T. and Cummins, T. K. (2003). Neurobiology of anxiety and depression. In Kasper, S., den Boer, J. A., and Sitsen, J. M. A., editors, *Handbook of Depression and Anxiety*, chapter 14, pages 127–49. Marcel Dekker, second edition. 23, 40

Niv, Y., Daw, N., and Dayan, P. (2005). How fast to work: Response vigor, motivation and tonic dopamine. In *Advances in Neural Information Processing*, pages 1019–26. MIT Press. 18, 45, 46, 104, 120, 122

Niv, Y., Daw, N. D., and Dayan, P. (2006). Choice values. *Nat Neurosci*, 9(8):987–988. 122

Niv, Y., Daw, N. D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, 191(3):507–520. 18, 28, 45, 46, 104, 120

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *J. Abnorm. Psychol.*, 100(4):569–82. 29, 116

Nouraei, R., Huys, Q. J. M., Chatrath, P., Powles, J., and Harcourt, J. (2007). Screening patients with sensorineural hearing loss for vestibular schwannoma using a bayesian classifier. *Clin. Otolaryng.*, In press. 11

Nutt, D. J. (2006). The role of dopamine and norepinephrine in depression and antidepressant treatment. *J Clin Psychiatry*, 67 Suppl 6:3–8. 27

O'Brien, J. T. (2006). Depression and comorbidity. *Am J Geriatr Psychiatry*, 14(3):187–190. 23

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669):452–4. 17

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, 38(2):329–337. 17

O'Doherty, J. P., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.*, 4(1):95–102. 17

Otter, C., Huber, J., and Bonner, A. (1995). Cloninger's tridimensional personality questionnaire: Reliability in an english sample. *Personality and Individual Differences*, 18(4):471–80. 107

Overmier, J. B., Patterson, J., and Wielkiewicz, R. M. (1980). Environmental contingencies as sources of stress in animals. In Levine, S. and Ursin, H., editors, *Coping and Health*. Plenum Press. 35, 54, 55, 84, 86, 105, 106, 137

Overmier, J. B. and Seligman, M. E. (1967). Effects of inescapable shock upon subsequent escape

and avoidance responding. *J Comp Physiol Psychol*, 63(1):28–33. 35, 53, 63, 86

Padesky, C. (1994). Schema change processes in cognitive therapy. *Clinical Psychology and Psychotherapy*, 1(5):267–278. 37

Papp, M., Klimek, V., and Willner, P. (1994). Parallel changes in dopamine D_2 receptor binding in limbic forebrain associated with chronic mild stress-induced anhedonia and its reversal by imipramine. *Psychopharmacology*, 115(4):441–6. 57

Papp, M., Lappas, S., Muscat, R., and Willner, P. (1992). Attenuation of place preference conditioning but not place aversion conditioning by chronic mild stress. *J. Psychopharmacol.*, 6:352–6. 55, 56

Papp, M., Muscat, R., and Willner, P. (1993a). Subsensitivity to rewarding and locomotor stimulant effects of a dopamine agonist following chronic mild stress. *Psychopharmacology*, 110(1-2):152–8. 57

Papp, M., Willner, P., and Muscat, R. (1993b). Behavioural sensitization to a dopamine agonist is associated with reversal of stress-induced anhedonia. *Psychopharmacology (Berlin)*, 110(1-2):159–64. 57

Parker, G. (2007). Defining melancholia: the primacy of psychomotor disturbance. *Acta Psychiatr Scand*, 115 (Suppl 433)(2):21–30. 44, 47

Parker, G. and Hadzi-Pavlovic, D. (1996). *Melancholia: A disorder of movement and mood*. Cambridge University Press. 18, 25, 44, 104

Parker, G., Hadzi-Pavlovic, D., Boyce, P., Wilhelm, K., Brodaty, H., Mitchell, P., Hickie, I., and Eyers, K. (1990). Classifying depression by mental state signs. *Br. J. Psychiatry*, 157:55–65. 44

Parker, G., Hadzi-Pavlovic, D., Wilhelm, K., Hickie, I., Brodaty, H., Boyce, P., Mitchell, P., and Eyers, K. (1994). Defining melancholia: properties of a refined sign-based measure. *Br. J. Psychiatry*, 164:316–26. 26

Parker, G. and Manicavasagar, V. (2005). *Modelling and managing the depressive disorders*. Cambridge University Press, Cambridge, UK. 26, 44

Parsey, R. V., Hastings, R. S., Oquendo, M. A., Hu, X., Goldman, D., yu Huang, Y., Simpson, N., Arcement, J., Huang, Y., Ogden, R. T., Heertum, R. L. V., Arango, V., and Mann, J. J. (2006a). Effect of a triallelic functional polymorphism of the serotonin-transporter-linked promoter region on expression of serotonin transporter in the human brain. *Am J Psychiatry*, 163(1):48–51. 34

Parsey, R. V., Hastings, R. S., Oquendo, M. A., yu Huang, Y., Simpson, N., Arcement, J., Huang, Y., Ogden, R. T., Heertum, R. L. V., Arango, V., and Mann, J. J. (2006b). Lower serotonin transporter binding potential in the human brain during major depressive episodes. *Am J Psychiatry*, 163(1):52–58. 34

Parsons, L. H. and Jr, J. B. J. (1993). Perfusate serotonin increases extracellular dopamine in the nucleus accumbens as measured by in vivo microdialysis. *Brain Res.*, 606(2):195–99. 60, 122

Peterson, C., Maier, S. F., and Seligman, M. E. P. (1993). *Learned Helplessness: A theory for the age of personal control*. OUP, Oxford, UK. 51, 63, 67, 81, 86

Peterson, C., Semmel, A., von Baeyer, C., Abramson, L., Metalsky, G., and Seligman, M. (1982). The attributional Style Questionnaire. *Cognitive Therapy and Research*, 6(3):287–299. 51

Pezawas, L., Meyer-Lindenberg, A., Drabant, E. M., Verchinski, B. A., Munoz, K. E., Kolachana, B. S., Egan, M. F., Mattay, V. S., and Weinberger, A. R. H. D. R. (2005). 5-httlpr polymorphism impacts human cingulate-amygdala interactions: a genetic susceptibility mechanism

for depression. *Nat. Neuosci.*, 8(6):828–34. 34

Phillips, A. G. and Barr, A. M. (1997). Effects of chronic mild stress on motivation for sucrose: mixed messages. *Psychopharmacology (Berl)*, 134(4):361–6; discussion 371–7. 55

Phillips, P. E. M., Stuber, G. D., Heien, M. L. A. V., Wightman, R. M., and Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature*, 422(6932):614–618. 18

Pilgrim, D. and Bentall, R. (1999). The medicalisation of misery: A critical realist analysis of the concept of depression. *J Mental Health*, 8(3):261–74. 25

Pizzagalli, D., Jahn, A., and OShea, J. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biol. Psychiatry*, 57(4):319–27. 17, 40, 41

Post, R. M., Kotin, J., Goodwin, F. K., and Gordon, E. K. (1973). Psychomotor activity and cerebrospinal fluid amine metabolites in affective illness. *Am J Psychiatry*, 130(1):67–72. 46

Power, M., editor (2005). *Mood Disorders: A Handbook of Science and Practice*. John Wiley and Sons, paperback edition. 23

Price, L. H., Charney, D. S., Delgado, P. L., and Heninger, G. R. (1991). Serotonin function and depression: neuroendocrine and mood responses to intravenous l-tryptophan in depressed patients and healthy comparison subjects. *Am J Psychiatry*, 148(11):1518–1525. 34

Puglisi-Allegra, S., Imperato, A., Angelucci, L., and Cabib, S. (1991). Acute stress induces time-dependent responses in dopamine mesolimbic system. *Brain Res.*, 554(1-2):217–21. 57

Purcell, R., Maruff, P., Kyrios, M., and Pantelis, C. (1997). Neuropsychological function in young patients with unipolar major depression. *Psychol Med*, 27(6):1277–1285. 42

Rachman, S. and Arntz, A. (1991). The overprediction and underprediction of pain. *Clinical Psychology Review*, 11:339–55. 116

Raghunathan, R. and Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. *Organ Behav Hum Decis Process*, 79(1):56–77. 46

Ranade, S. P. and Mainen, Z. F. (2006). Tetrode recordings in dorsal and median raphe nuclei in awake behaving rats. In *Comp. Sys. Neurosci.* 19

Randrup, A., Munkvad, I., Fog, R., Gerlach, J., Molander, L., Kjellberg, B., and Scheel-Kruger, J. (1975). Mania, depression and brain dopamine. *Current developments in psychopharmacology*, 2:206–248. 27, 47

Rang, H. P., Dale, M. M., and Ritter, J. M. (2000). *Pharmacology*. Churchill Livingstone, Edinburgh, UK, 4th edition. 33

Redgrave, P. (1978). Modulation of intracranial self-stimulation behaviour by local perfusions of dopamine, noradrenaline and serotonin within the caudate nucleusand nucleus accumbens. *Brain Res.*, 155(2):277–95. 60

Rescorla, R. and Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, pages 64–99. 17

Ressler, K. J. and Nemeroff, C. B. (2000). Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders. *Depression Anxiety*, 12(S1):2–19. 122

Richards, P. M. and Ruff, R. M. (1989). Motivational effects on neuropsychological functioning: comparison of depressed versus nondepressed individuals. *J Consult Clin Psychol*, 57(3):396–402. 17, 45

Richell, R. A., Deakin, J. F. W., and Anderson, I. M. (2005). Effect of acute tryptophan depletion

on the response to controllable and uncontrollable noise stress. *Biol Psychiatry*, 57(3):295–300. 34

Rizley, R. (1978). Depression and distortion in the attribution of causality. *J Abnorm Psychol*, 87(1):32–48. 37

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., and Rabbitt, P. (1994). Cambridge neuropsychological test automated battery (cantab): a factor analytic study of a large sample of normal elderly volunteers. *Dementia*, 5(5):266–281. 26

Robertson, M. M. and Trimble, M. R. (1981). Neuroleptics as antidepressants. *Neuropharmacology*, 20(12B):1335–1336. 47

Robertson, M. M. and Trimble, M. R. (1982). Major tranquillisers used as antidepressants. a review. *J Affect Disord*, 4(3):173–193. 47

Robinson, O. J., Cools, R., and Sahakian, B. J. (2007). Enhanced punishment but not reward prediction under tryptophan depletion. In *Depression. Brain causes – body consequences*, London, UK. Institute of Pyschiatry conference poster. 33, 34

Rocha, B. A., Goulding, E. H., O'Dell, L. E., Mead, A. N., Coufal, N. G., Parsons, L. H., and Tecott, L. H. (2002). Enhanced locomotor, reinforcing, and neurochemical effects of cocaine in serotonin 5-hydroxytryptamine 2c receptor mutant mice. *J Neurosci*, 22(22):10039–10045. 59

Rocha, B. A., Scearce-Levie, K., Lucas, J. J., Hiroi, N., Castanon, N., Crabbe, J. C., Nestler, E. J., and Hen, R. (1998). Increased vulnerability to cocaine in mice lacking the serotonin-1b receptor. *Nature*, 393(6681):175–8. 59

Rogers, M. A., Bellgrove, M. A., Chiu, E., Mileshkin, C., and Bradshaw, J. L. (2004). Response selection deficits in melancholic but not nonmelancholic unipolar major depression. *J. Clin. Exp. Neuropsychology*, 26(2):169–79. 44

Rogers, M. A., Bradshaw, J. L., Phillips, J. G., Chiu, E., Vaddadi, K., Presnel, I., and Mileshkin, C. (2000). Parkinsonian motor characteristics in unipolar major depression. *J. Clin. Exp. Neuropsychology*, 22(2):232–44. 47

Rogers, R. D., Blackshaw, A. J., Middleton, H. C., Matthews, K., Hawtin, K., Crowley, C., Hopwood, A., Wallace, C., Deakin, J. F. W., Sahakian, B. J., and Robbins, T. W. (1999). Tryptophan depletion impairs stimulus-reward learning while methylphenidate disrupts attentional control in healthy young adults: implication for the monoaminergic basis of impulsive behaviour. *Psychopharm.*, 146:428–91. 43, 122

Rogers, R. D., Tunbridge, E. M., Bhagwagar, Z., Drevets, W. C., Sahakian, B. J., and Carter, C. S. (2003). Tryptophan depletion alters the decision-making of healthy volunteers through altered processing of reward cues. *Neuropsychopharmacology*, 28(1):153–62. 34, 122

Roiser, J. P., Blackwell, A. D., Cools, R., Clark, L., Rubinsztein, D. C., Robbins, T. W., and Sahakian, B. J. (2006). Serotonin transporter polymorphism mediates vulnerability to loss of incentive motivation following acute tryptophan depletion. *Neuropsychopharmacology*, 31(10):2264–2272. 34, 121, 122

Rosellini, R. (1978). Inescapable shock interferes with acquisition of an appetitive operant. *Animal Learning and Behavior*, 6(2):155–159. 54

Rosellini, R. A., DeCola, J. P., and Shapiro, N. R. (1982). Cross-motivational effects of inescapable shock are associative in nature. *J Exp Psychol Anim Behav Process*, 8(4):376–388. 54

Roth, S. and Bootzin, R. (1974). Effects of experimentally induced expectancies of external control: An investigation of learned helplessness. *Journal of Personality and Social Psychology*, 29:253–264. 106

Roth, S. and Kubal, L. (1975). Effects of noncontingent reinforcement on tasks of differing importance: Facilitation and learned helplessness. *Journal of Personality and Social Psychology*, 32:680–691. 36, 63, 106

Rottenberg, J., Kasch, K. L., Gross, J. J., and Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion*, 2(2):135–46. 29, 32, 105

Ruchsow, M., Herrnberger, B., Beschoner, P., Grön, G., Spitzer, M., and Kiefer, M. (2005). Error processing in major depressive disorder: Evidence from event-related potentials. *J. Psychiatr. Res.*, 40(1):37–46. 42

Ruchsow, M., Herrnberger, B., Wiesend, C., Gron, G., Spitzer, M., and Kiefer, M. (2004). The effect of erroneous responses on response monitoring in patients with major depressive disorder: a study with event-related potentials. *Psychophysiology*, 41(6):833–40. 42

Sachdev, P. and Aniss, A. M. (1994). Slowness of movement in melancholic depression. *Biol. Psychiatry*, 35(4):253–62. 47

Salamone, J. D. and Correa, M. (2002). Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behav Brain Res*, 137(1-2):3–25. 18

Santarelli, L., Saxe, M., Gross, C., Surget, A., Battaglia, F., Dulawa, S., Weisstaub, N., Lee, J., Duman, R., Arancio, O., Belzung, C., and Hen, R. (2003). Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science*, 201:805–9. 20, 127

Sapolsky, R. M. (2004). *Why zebras don't get ulcers*. Henry Holt and Co., New York, USA, third edition. 30

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science*, 308:648–52. 30

Sarek, M. (2006). Evident exception in clinical practice not sufficient to break traditional hypothesis. *PLoS Med*, 3(2):e120; author reply e116. 33

Sasaki-Adams, D. M. and Kelley, A. E. (2001). Serotonin-dopamine interactions in the control of conditioned reinforcement and motor behaviour. *Neuropsychopharm*, 25(3):440–52. 60

Scheggi, S., Leggio, B., Masi, F., Grappi, S., Gambarana, C., Nanni, G., Rauggi, R., and De Montis, M. G. (2002). Selective modifications in the nucleus accumbens of dopamine synaptic transmission in rats exposed to chronic stress. *J. Neurochem.*, 83(4):893–903. 57

Schmajuk, N. A., Gray, J. A., and Lam, Y. W. (1996). Latent inhibition: a neural network approach. *J Exp Psychol Anim Behav Process*, 22(3):321–349. 110

Schmand, B., et al., and et al. (1994). Cognitive disorders and negative symptoms as correlates of motivational deficits in psychotic patients. *Psychol. Med.*, 26:869–84. 43

Schruers, K., Klaassen, T., Pols, H., Overbeek, T., Deutz, N. E., and Griez, E. (2000). Effects of tryptophan depletion on carbon dioxide provoked panic in panic disorder patients. *Psychiatry Res*, 93(3):179–187. 35

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J Neurophysiol*, 80(1):1–27. 17, 82

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward.

Science, 275(5306):1593–1599. 17, 82, 122

Schultz, W. and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annu Rev Neurosci*, 23:473–500. 19

Schweighofer, N., Shishida, K., Han, C. E., Okamoto, Y., Tanaka, S. C., Yamawaki, S., and Doya, K. (2006). Humans can adopt optimal discounting strategy under real-time constraints. *PLoS Comput Biol*, 2(11):e152. 34, 121, 122

Schweighofer, N., Tanaka, S. C., and Doya, K. (2007). Serotonin and the evaluation of future rewards: Theory, experiments, and possible neural mechanisms. *Ann N Y Acad Sci*. 34, 121, 122

Seligman, M. E. and Maier, S. F. (1967). Failure to escape traumatic shock. *J Exp Psychol*, 74(1):1–9. 35, 53, 66, 86, 108

Seligman, M. E. P. (1975). *Helplessness. On Depression, Development and Death*. W. H. Freeman & Co., San Francisco, USA. 14, 53, 77, 86, 108

Selye, H. (1984). *The Stress of Life*. McGraw-Hill. 30

Serra, G., Argiolas, A., Klimek, V., Fadda, F., and Gessa, G. L. (1979). Chronic treatment with antidepressants prevents the inhibitory effect of small doses of apomorphine on dopamine synthesis and motor activity. *Life Sci*, 25(5):415–423. 60

Shah, P. J., O'Carroll, R. E., Rogers, A., Moffoot, A. P., and Ebmeier, K. P. (1999). Abnormal response to negative feedback in depression. *Psychol Med*, 29(1):63–72. 42

Shah, P. J., Ogilvie, A. D., Goodwin, G. M., and Ebmeier, K. P. (1997). Clinical and psychometric correlates of dopamine d2 binding in depression. *Psychol Med*, 27(6):1247–1256. 46

Shallice, T. (1982). Specific impairments of planning. *Philos Trans R Soc Lond B Biol Sci*, 298(1089):199–209. 38

Shanks, D. R. (1995). *The psychology of associative Learning*. Number 13 in Problems in Behavioural Sciences. Cambridge University Press, Cambridge, UK. 134

Sibille, E. and Lewis, D. A. (2006). Sert-ainly involved in depression – but when? *Am J Psychiatry*, 163:8–11. 32, 33

Simpson, S., Corney, R., Fitzgerald, P., and Beecham, J. (2003). A randomized controlled trial to evaluate the effectiveness and cost-effectiveness of psychodynamic counselling for general practice patients with chronic depression. *Psychol Med*, 33(2):229–239. 27

Sims, A. C. P. (2003). *Symptoms of the Mind: And introduction to descriptive psychopathology*. Saunders, 3rd edition. 28

Smith, A., Li, M., Becker, S., and Kapur, S. (2004). A model of antipsychotic action in conditioned avoidance: a computational approach. *Neuropsychopharm.*, 29(6):1040–9. 39, 103

Smith, A., Li, M., Becker, S., and Kapur, S. (2006). Dopamine, prediction error and associative learning: a model-based account. *Network*, 17(1):61–84. 103

Smith, A. J., Becker, S., and Kapur, S. (2005). A computational model of the functional role of the ventral-striatal d2 receptor in the expression of previously acquired behaviors. *Neural Comput*, 17(2):361–95. 103

Smith, K. A., Fairburn, C. G., and Cowen, P. J. (1997). Relapse of depression after rapid depletion of tryptophan. *Lancet*, 249:915–9. 33, 34, 116, 121

Smith, K. A., Fairburn, C. G., and Cowen, P. J. (1999). Symptomatic relapse in bulimia nervosa following acute tryptophan depletion. *Arch. Gen. Psych.*, 56:171–6. 32, 110, 121

Solomon, R. L. and Corbit, J. D. (1974). An opponent-process theory of motivation. i. temporal

dynamics of affect. *Psychol Rev*, 81(2):119–145. 110

Soubrié, P. (1986). Reconciling the role of central serotonin neurons in human and animal behaviour. *Behav Brain Sci*, 9:319–364. 19, 32, 59, 110

Spitzer, L., Endicott, J., and Robins, E. (1978). Research Diagnostic Criteria (RDC). Biometric Research. 25

Spitzer, R. L. (1998). Diagnosis and need for treatment are not the same. *Arch Gen Psychiatry*, 55(2):120. 24, 26

Stamford, J. A., Muscat, R., O'Connor, J. J., Patel, J., Trout, S. J., Wieczorek, W. J., Kruk, Z. L., and Willner, P. (1991). Voltammetric evidence that subsensitivity to reward following chronic mild stress is associated with increased release of mesolimbic dopamine. *Psychopharmacology*, 105(2):275–81. 57, 83

Steffens, D. C., Wagner, H. R., Levy, R. M., Horn, K. A., and Krishnan, K. R. (2001). Performance feedback deficit in geriatric depression. *Biol Psychiatry*, 50(5):358–363. 41

Stevens, A. and Price, J. (2000). *Evolutionary Psychiatry. A New Beginning*. Routledge, London, UK, second edition. 103, 127

Strekalova, T., Spanagel, R., Bartsch, D., Henn, F. A., and Gass, P. (2004). Stress-induced anhedonia in mice is associated with deficits in forced swimming and exploration. *Neuropsychopharmacol.*, 29(11):2007–11. 52, 55, 59, 107, 108

Strens, M. (2000). A bayesian framework for reinforcement learning. In *Proceeedings of the 17th International Conference on Machine Learning (ICML)*. 104

Strickland, P. L., Deakin, J. F. W., Percival, C., Dixon, J., Gater, R. A., and Goldberg, D. P. (2002). Bio-social origins of depression in the community: Interactions between social adversity, cortisol and serotonin neurotransmission. *Br. J. Psychiatry*, 180:168–73. 30

Ströhle, A. and Holsboer, F. (2003). Stress-responsive neurohormones in depression and anxiety. In Kasper, S., den Boer, J. A., and Sitsen, J. M. A., editors, *Handbook of Depression and Anxiety*, chapter 10, pages 207–228. Marcel Dekker, second edition. 30

Styron, W. (1991). *Darkness Visible*. Jonathan Cape. 14

Sulzer, D. and Edwards, R. H. (2005). Antidepressants and the monoamine masquerade. *Neuron*, 46(1):1–2. 60

Sutton, L. C., Grahn, R. E., Wiertelak, E. P., Watkins, L. R., and Maier, S. F. (1997). Inescapable shock-induced potentiation of morphine analgesia in rats: involvement of opioid, gabaergic, and serotonergic mechanisms in the dorsal raphe nucleus. *Behav Neurosci*, 111(4):816–824. 58, 81

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA. 17, 63, 64, 67, 83, 86, 111, 113, 114, 127, 134

Sweeney, P. D., Anderson, K., and Bailey, S. (1986). Attributional style in depression: a meta-analytic review. *J Pers Soc Psychol*, 50(5):974–991. 37

Szabadi, E., Bradshaw, C. M., and Ruddle, H. V. (1981). Reinforcement processes in affective illness: towards a quantitative measure. In Bradshaw, C. M., Szabadi, E., and Lowe, C. F., editors, *Quantification of steady-state operant behavior*, pages 299–310. Elsevier/North-Holland Biomedical Press. 45, 46

Takase, L. F., Nogueira, M. I., Baratta, M., Bland, S. T., Watkins, L. R., Maier, S. F., Fornal, C. A., and Jacobs, B. L. (2004). Inescapable shock activates serotonergic neurons in all raphe nuclei of rat. *Behav Brain Res*, 153(1):233–239. 32, 58, 82

Takase, L. F., Nogueira, M. I., Bland, S. T., Baratta, M., Watkins, L. R., Maier, S. F., Fornal, C. A., and Jacobs, B. L. (2005). Effect of number of tailshocks on learned helplessness and activation of serotonergic and noradrenergic neurons in the rat. *Behav Brain Res*, 162(2):299–306. 58, 82

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., and Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.*, 7(8):887–93. 122

Terman, G. W., Shavit, Y., Lewis, J. W., Cannon, J. T., and Liebeskind, J. C. (1984). Intrinsic mechanisms of pain inhibition: activation by stress. *Science*, 226(2680):1270–7. 54, 63, 81

Testa, T. J., Juraska, J. M., and Maier, S. F. (1974). Prior exposure to inescapable electric shock in rats affects extinction behavior after the successful acquisition of an escape response. *Learning Memory*, 5(3):380–92. 53

Träskman, L., Asberg, M., Bertilsson, L., and Sjöstrand, L. (1981). Monoamine metabolites in csf and suicidal behavior. *Arch Gen Psychiatry*, 38(6):631–636. 33

Tremblay, L. K., Naranjo, C. A., Cardenas, L., Herrmann, N., and Busto, U. E. (2002). Probing brain reward system function in major depressive disorder: altered response to dextroamphetamine. *Arch. Gen. Psych.*, 59(5):209–15. 34, 47

Tremblay, L. K., Naranjo, C. A., Graham, S. J., Herrmann, N., Mayberg, H. S., Hevenor, S., and Busto, U. E. (2005). Functional neuroanatomical substrates of altered reward processing in major depressive disorder revealed by a dopaminergic probe. *Arch. Gen. Psych.*, 62(11):1228–36. 47

Trompenaars, F. J., Masthoff, E. D., Heck, G. L. V., Hodiamont, P. P., and Vries, J. D. (2006). Relationship between mood related disorders and quality of life in a population of dutch adult psychiatric outpatients. *Depress Anxiety*, 23(6):353–363. 23

Ungless, M. A. (2004). Dopamine: the salient issue. *Trends Neurosci.*, 27(12):702–6. 56

Ungless, M. A., Magill, P. J., and Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science*, 303(5666):2040–2. 56

Unterrainer, J. M., Kaller, C. P., Halsband, U., and Rahm, B. (2006). Planning abilities and chess: a comparison of chess and non-chess players on the tower of london task. *Br J Psychol*, 97(Pt 3):299–311. 38

Van Os, J., Gilvarry, C., Bale, R., Horn, E. V., Tattan, T., White, I., and Murray, R. (1999). A comparison of the utility of dimensional and categorical representations of psychosis. uk700 group. *Psychol Med*, 29(3):595–606. 25

Veiel, H. O. F. (1997). A preliminary profile of neuropsychological deficits associated with major depression. *J. Clin. Exp. Neuropsychol.*, 19:587–603. 43, 44

Velten, E. (1968). A laboratory task for induction of mood states. *Behav Res Ther*, 6(4):473–482. 45

Volpicelli, J., Ulm, R., Altenor, A., and Seligman, M. (1983). Learned mastery in the rat. *Learning and Motivation*, 14:204–222. 53, 54

Vossen, H. G. M., van Os, J., Hermens, H., and Lousberg, R. (2006). Evidence that trait-anxiety and trait-depression differentially moderate cortical processing of pain. *Clin J Pain*, 22(8):725–729. 31

Vowles, K. E., McNeil, D. W., Sorrell, J. T., , and Lawrence, S. M. (2006). Fear and pain: Investigating the interaction between aversive states. *J Abnormal Psych*, 115(4):821–33. 54

Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic as-

sumptions of formal learning theory. *Nature*, 412(6842):43–48. 17

Walsh, M.-T. and Dinan, T. G. (2001). Selective serotonin reuptake inhibitors and violence: a review of the available evidence. *Acta Psychiatr Scand*, 104(2):84–91. 121

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., and Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *J Exp Psych: Learn Mem Cog*, 19(1):1. 38

Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292. 17, 67, 135

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., and McCormick, R. A. (1995). Testing a tripartite model: I. evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J Abnorm Psychol*, 104(1):3–144. 43

Weiss, J. M., Bailey, W. H., Pohorecky, L. A., Korzeniowski, D., and Grillione, G. (1980). Stress-induced depression of motor activity correlates with regional changes in brain norepinephrine but not in dopamine. *Neurochem Res*, 5(1):9–22. 35

Weiss, J. M., Goodman, P. A., Losito, B. G., Corrigan, S., Charry, J. M., and Bailey, W. H. (1981). Behavioral depression produced by an uncontrollable stressor: Relationship to norepinephrine, dopamine, and serotonin levels in various regions of rat brain. *Brain Res Rev*, 3(2):167–205. 35, 63

Weissman, A. and Beck, A. (1978). Development and validation of the Dysfunctional Attitude Scale. *Annual Meeting of the Association for the Advancement of Behavior Therapy, Chicago*. 51

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H. G., Joyce, P. R., Karam, E. G., Lee, C. K., Lellouch, J., Lpine, J. P., Newman, S. C., Rubio-Stipec, M., Wells, J. E., Wickramaratne, P. J., Wittchen, H., and Yeh, E. K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *JAMA*, 276(4):293–299. 24

Weissman, M. M., McAvay, G., Goldstein, R. B., Nunes, E. V., Verdeli, H., and Wickramaratne, P. J. (1999a). Risk/protective factors among addicted mothers' offspring: a replication study. *Am J Drug Alcohol Abuse*, 25(4):661–679. 24

Weissman, M. M., Wolk, S., Wickramaratne, P., Goldstein, R. B., Adams, P., Greenwald, S., Ryan, N. D., Dahl, R. E., and Steinberg, D. (1999b). Children with prepubertal-onset major depressive disorder and anxiety grown up. *Arch Gen Psychiatry*, 56(9):794–801. 24

Weisstaub, N. V., Zhou, M., Lira, A., Lambe, E., Gonzlez-Maeso, J., Hornung, J.-P., Sibille, E., Underwood, M., Itohara, S., Dauer, W. T., Ansorge, M. S., Morelli, E., Mann, J. J., Toth, M., Aghajanian, G., Sealfon, S. C., Hen, R., and Gingrich, J. A. (2006). Cortical 5-ht2a receptor signaling modulates anxiety-like behaviors in mice. *Science*, 313(5786):536–540. 59

Welker, R. (1976). Acquisition of a free-operant-appetitive response in pigeons as a function of prior experience with response-independent food. *Learning and Motivation*, 7:394–405. 54

Wener, A. E. and Rehm, L. P. (1975). Depressive affect: a test of behavioral hypotheses. *J. Abnorm. Psychol.*, 84(3):221–7. 29

Whale, R., Quested, D. J., Laver, D., Harrison, P. J., and Cowen, P. J. (2000). Serotonin transporter (5-htt) promoter genotype may influence the prolactin response to clomipramine. *Psychopharmacology (Berl)*, 150(1):120–122. 32

Wilhelm, K., Mitchell, P. B., Niven, H., Finch, A., Wedgwood, L., Scimone, A., Blair, I. P., Parker, G., and Schofield, P. R. (2006). Life events, first depression onset and the serotonin transporter gene. *Br J Psychiatry*, 188:210–5. 50

Williams, J. and Dayan, P. (2005). Dopamine, learning, and impulsivity: a biological account

of attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol*, 15(2):160–79; discussion 157–9. 103

Williams, J. M. G. (1992). *The psychological treatment of depression*. Routledge. 29, 50, 63

Willis, M. H. and Blaney, P. H. (1978). Three tests of the learned helplessness model of depression. *J Abnorm Psychol*, 87(1):131–136. 36

Willner, P. (1983a). Dopamine and depression: a review of recent evidence. I. Empirical studies. *Brain Res. Rev.*, 287(3):211–24. 46

Willner, P. (1983b). Dopamine and depression: a review of recent evidence. II. Theoretical approaches. *Brain Res. Rev.*, 287(3):225–36. 53

Willner, P. (1985a). Antidepressants and serotonergic neurotransmission: an integrative review. *Psychopharmacology (Berl)*, 85(4):387–404. 59

Willner, P. (1985b). *Depression: A psychobiological synthesis*. John Wiley & Sons, New York. 19, 27, 28, 32, 33, 34, 46, 47, 52, 62, 63, 110, 123

Willner, P. (1986). Validation criteria for animal models of human mental disorders: learned helplessness as a paradigm case. *Prog Neuropsychopharmacol Biol Psychiatry*, 10(6):677–690. 52, 62, 86

Willner, P. (1991). Animal models as simulations of depression. *Trends Pharmacol. Sci.*, 12(4):131–6. 83

Willner, P. (1997). Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. *Psychopharm*, 134:319–29. 28, 52, 55, 86, 89, 100, 105

Willner, P. (2002). Dopamine and depression. In Chiara, G. D., editor, *Handbook of Physiology: Dopamine in the CNS.*, pages 387–416. Springer, Berlin. 18, 28, 46, 47, 123

Willner, P. (2005). Chronic mild stress (cms) revisited: Consistency and behavioural-neurobiological concordance in the effects of cms. *Neuropsychobiology*, 52(2):90–110. 55, 56

Willner, P., Benton, D., Brown, E., Cheeta, S., Davies, G., Morgan, J., and Morgan, M. (1998). "depression" increases "craving" for sweet rewards in animal and human models of depression and craving. *Psychopharmacology*, 136(3):272–83. 46, 55

Willner, P., Hale, A. S., and Argyropoulos, S. (2005). Dopaminergic mechanism of antidepressant action in depressed patients. *J. Affect. Disord.*, 86(1):37–45. 47, 60, 122

Willner, P. and Healy, S. (1994). Decreased hedonic responsiveness during a brief depressive mood swing. *J. Affect. Disord.*, 32(1):13–20. 46

Willner, P. and Jones, C. (1996). Effects of mood manipulation on subjective and behavioural measures of cigarette craving. *Behav Pharmacol*, 7(4):355–363. 46

Willner, P., Klimek, V., Golembiowska, K., and Muscat, R. (1991). Changes in mesolimbic dopamine may explain stress-induced anhedonia. *Psychobiology*, 19:79–84. 57

Willner, P. and Mitchell, P. J. (2002). The validity of animal models of predisposition to depression. *Behav. Pharmacol.*, 13(3):169–88. 52

Willner, P. and Mitchell, P. J. (2003). Animal models of subtypes of depression. In Kasper, S., den Boer, J. A., and Sitsen, J. M. A., editors, *Handbook of Depression and Anxiety*, chapter 2, pages 505–44. Marcel Dekker, second edition. 52, 55, 59, 62, 86

Willner, P., Muscat, R., and Papp, M. (1992a). Chronic mild stress-induced anhedonia: A realistic animal model of depression. *Neurosci. Biobehav. Rev.*, 16(4):525–34. 83

Willner, P., Phillips, G., Muscat, R., and Hood, P. (1992b). Behavioural tests of the dopamine depletion hypothesis of neuroleptic-induced response decrement. *Psychopharmacology*,

106(4):543. 83

Willner, P., Towell, A., Sampson, D., Sophokleous, S., and Muscat, R. (1987). Reduction of sucrose preference by chronic unpredictable mild stress, and its restoration by a tricyclic antidepressant. *Psychopharmacology*, 93(3):358–64. 28, 55, 89, 108

Wise, C. D., Berger, B. D., and Stein, L. (1972). Benzodiazepines: Anxiety-reducing activity by reduction of serotonin turnover in the brain. *Science*, 177(4044):180–3. 19

Wolff, E. A., Putnam, F. W., and Post, R. M. (1985). Motor activity and affective illness. the relationship of amplitude and temporal distribution to changes in affective state. *Arch Gen Psychiatry*, 42(3):288–294. 44

Wong, M.-L. and Licinio, J. (2001). Research and treatment approaches to depression. *Nat. Neurosci. Rev.*, 2:343. 13, 23, 24, 62

Woolfe, R., Dryden, W., and Strawbridge, S., editors (2003). *Handbook of counselling psychology*. Sage, second edition. 63

World Health Organization (1990). *International Classification of Diseases*. World Health Organization Press. 25

World Health Organization (1996). *The global burden of disease*. World Health Organization Press. 23

Wortman, C. and Brehm, J. (1975). Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. *Advances in experimental social psychology*, 8(S 278):336. 81, 106

Young, S. N., Smith, S. E., Pihl, R. O., and Ervin, F. R. (1985). Tryptophan depletion causes a rapid lowering of mood in normal males. *Psychopharmacology (Berl)*, 87(2):173–177. 33, 110, 121

Yu, Y. W.-Y., Tsai, S.-J., Chen, T.-J., Lin, C.-H., and Hong, C.-J. (2002). Association study of the serotonin transporter promoter polymorphism and symptomatology and antidepressant response in major depressive disorders. *Mol Psychiatry*, 7(10):1115–1119. 32

Zacharko, R. M. and Anisman, H. (1991). Stressor-induced anhedonia in the mesocorticolimbic system. *Neurosci. Biobehav. Rev.*, 15(3):391–405. 57, 105

Zacharko, R. M., Bowers, W. J., Kokkinidis, L., and Anisman, H. (1983). Region-specific reductions of intracranial self-stimulation after uncontrollable stress: possible effects on reward processes. *Behav. Brain Res.*, 9(2):129–41. 54, 57, 83, 105

Zemel, R. S., Huys, Q. J. M., Natarajan, R., and Dayan, P. (2005). Probabilistic computation in spiking populations. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems (NIPS) 17*, pages 1609–1616. MIT Press, Cambridge, MA. 11

Zhao, Z.-Q., Scott, M., Chiechio, S., Wang, J.-S., Renner, K. J., Gereau, R. W., Johnson, R. L., Deneris, E. S., and Chen, Z.-F. (2006). Lmx1b is required for maintenance of central serotonergic neurons and mice lacking central serotonergic system exhibit normal locomotor activity. *J Neurosci*, 26(49):12781–12788. 59